

PIECEWISE GOMPERTZ MODEL ON SOLVING CURE RATE PROBLEM



A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2001

Copyright 2001

by

Haoyi Chen

I dedicate this work to Hua.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest gratitude to Professor Myron N. Chang, my advisor, for his guidance, tremendous help, and encouragement during my doctoral research. This dissertation would not have been possible without his forceful guidance and excellent advisement.

I also would like to thank Professor Randolph R. Carter, my graduate research supervisor, for his guidance and support during the various epidemiological research projects with which I have been involved as a graduate research assistant. My deepest appreciation is extended to the other committee members, Dr. Ronald Randles, Dr. Samuel S. Wu, and Dr. Michael B. Resnick, for their advice on my doctoral research.

Lastly, I would like to extend my gratitude to my husband Hua and my parents for their never-ending support, understanding, and love.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	iv
ABSTRACT	vii
CHAPTERS	
1 INTRODUCTION	1
1.1 Problem and Common Practice	1
1.1.1 Mixture Model	2
1.1.2 Binary Regression Model	6
1.1.3 Nonparametric Method	7
1.1.4 Model with Bounded Cumulative Hazard Rate	10
1.2 Computation Difficulty of the Semiparametric Model	17
1.3 Piecewise Exponential Model	20
2 TOPIC OF RESEARCH	25
2.1 Data Structure and Model	26
2.2 Assumptions	28
2.3 Hypothesis of Interest	29
2.4 Test of Hypothesis	29
3 ESTIMATION OF PARAMETERS	31
3.1 Likelihood Function and Information Matrices	31
3.2 Concavity of Likelihood Function	32
4 PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATES	35
4.1 Introduction	35
4.2 Preliminary Proofs	36
4.3 Existence and Consistency of the MLE	37
4.4 Asymptotic Normality of the MLE	40
4.5 Goodness-of-Fit Test	42
5 SIMULATIONS	47
5.1 Introduction	47
5.2 Generation of Random Samples	47
5.2.1 Parameter Settings	48

5.2.2	Generation of Censored Observations	50
5.2.3	Generation of Failure Observations	56
5.3	Type I Error and Power	57
6	EXAMPLE	76
6.1	Introduction	76
6.2	Methodology	77
6.3	Results	80
6.3.1	Pairwise Comparisons on Cure Proportions	80
6.3.2	Cure Proportion Estimates	82
6.3.3	Goodness-of-Fit	82
7	SUMMARY	88
APPENDICES		
A	DERIVATION OF MARGINAL LIKELIHOOD	89
B	PROOF OF LEMMA 4.1	94
C	PROOF OF LEMMA 4.2	97
D	PROOF OF LEMMA 4.3	100
E	PROOF OF LEMMA 4.4	101
F	PROOF OF LEMMA 4.5	107
G	PROOF OF LEMMA 4.6	109
H	PROOF OF LEMMA 4.7	112
I	PROOF OF LEMMA 4.8	113
J	SIMPSON'S FORMULA	114
REFERENCES		115
BIOGRAPHICAL SKETCH		118

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

PIECEWISE GOMPERTZ MODEL ON SOLVING CURE RATE PROBLEM

By

Haoyi Chen

December 2001

Chair: Myron N. Chang
Major Department: Statistics

In cancer research, some of the patients are disease-free after certain treatments. It is interesting to compare treatment efficacy with long-term survival rates. One commonly used approach in analyzing this type of data is to compare the Kaplan-Meier estimates of the cure rate. Another approach is to apply the mixture model proposed by Farewell. However, the Kaplan-Meier estimates are unstable toward the end point, while the mixture model is computationally too complex.

To overcome these difficulties, we propose a test based on a piecewise Gompertz model to compare drug efficacy on the cure rate. The proposed test also accommodates the situation where patients display different hazard patterns during different treatment stages. In this work, we have derived the strict concavity of the log-likelihood function and the existence, consistency, and asymptotic normality of the maximum likelihood estimates of the parameters. In addition, our Monte Carlo simulation study shows that the proposed test is more computationally feasible and more powerful than the test based on Farewell's mixture model. Moreover, an

example is given to show the utility of our proposed test. A goodness-of-fit test of our proposed model is also discussed.

CHAPTER 1 INTRODUCTION

1.1 Problem and Common Practice

In survival analysis, one is interested in comparing the efficiency of various drugs or treatments. The survival probability is generally assumed to converge to zero when the observation time becomes large. However, it is not uncommon in cancer research that a proportion of patients are disease-free (relapse-free) after certain treatments. These patients are called “long-term survivors” or “cured.” For instance, in clinical trials dealing with childhood leukemia, some treatments do cure a proportion of patients permanently. This type of long-term survival pattern is different from what is dealt with under the traditional survival analysis. In addition, it is sometimes necessary to assume that hazard functions are different on different intervals. A piecewise exponential model proposed by Friedman (1981) accommodates this assumption. This model was considered semiparametric by Aitkin et al. (1989, PP. 287-292). They stated that the piecewise exponential model has the same flexibility as the proportional hazards model in which the baseline hazard function is unspecified. In this chapter, we first introduce the common methodologies in the literature that deal with cure rate problems. We also discuss the piecewise exponential model proposed by Friedman which assumes different hazard functions on different time intervals.

In the literature, there are several methods dealing with the long-term survival problem. They are as follows:

- Using a mixture model which assumes the underlying population is a mixture of two populations, patients who are cured and patients who are not cured.

- Applying a modified binary regression model in which the main interest is to compare the cure rates among patient groups.
- Using nonparametric methods to compare the survival curves or to compare cure proportions among treatment arms.
- Assuming a bounded cumulative hazard rate instead of a cumulative hazard rate with an infinite limit.

In the next four sections, we describe each method in more detail. Assume that there are n patients in the study. For the i^{th} patient, let T_i denote the time to relapse (incidence), U_i the censoring time, $T_i^* = \min(T_i, U_i)$, $\Delta_i = I_{[T_i \leq U_i]}$, and $\mathbf{x}_i$ a p -dimensional covariates vector associated with the patient. The observed data are $(t_1^*, \delta_1, \mathbf{x}_1), (t_2^*, \delta_2, \mathbf{x}_2), \dots, (t_n^*, \delta_n, \mathbf{x}_n)$.

1.1.1.1 Mixture Model

Let Z be a binary variable showing that a patient belongs to either the cured population or the noncured population.

$$Z = \begin{cases} 1, & \text{if the patient is noncured;} \\ 0, & \text{if the patient is cured.} \end{cases}$$

Note that Z is not fully observable in the presence of censoring. Clearly, if the incidence is observed, then $\Delta = 1$ and $Z = 1$; if the time to incidence is censored, then $\Delta = 0$ and Z is not observable. Assume that the probability density function for the time to incidence is $f^*(t \mid Z = 1)$ for the noncured population. Let $S^*(t \mid Z = 1)$ be the corresponding survival function. If $\delta_i = 1$, then the contribution of the i^{th} patient to the likelihood is

$$f^*(t_i \mid Z_i = 1)P(Z_i = 1).$$

If $\delta_i = 0$, then the contribution of the patient to the likelihood is

$$P(Z_i = 0) + P(Z_i = 1)S^*(t_i \mid Z_i = 1). \quad (1.1)$$

Therefore, the overall likelihood function is

$$L = \prod_{i=1}^n \{P(Z_i = 1)f^*(t_i^* | Z_i = 1)\}^{\delta_i} \{P(Z_i = 0) + P(Z_i = 1)S^*(t_i^* | Z_i = 1)\}^{1-\delta_i}. \quad (1.2)$$

Two influential articles addressing this method were written by Farewell in 1977 and 1982. In the first article, he expressed the binary variable Z using a logistic regression model

$$P(Z = 1) = \frac{e^{\beta^T \mathbf{x}}}{1 + e^{\beta^T \mathbf{x}}}, \quad (1.3)$$

where β is a vector of coefficients associated with the vector of covariates $\mathbf{x}$.

Farewell (1977) also assumed the time to incidence T follows an exponential distribution with mean $1/\lambda$, and λ is assumed to be known. Thus, when $\delta_i = 1$, the contribution of the i^{th} patient to the likelihood is

$$f^*(T_i = t_i; \lambda | Z_i = 1)P(Z_i = 1) = \lambda e^{-\lambda t_i} \frac{e^{\beta^T \mathbf{x}_i}}{1 + e^{\beta^T \mathbf{x}_i}}, \quad t_i \geq 0.$$

When $\delta_i = 0$, the contribution of the i^{th} patient to the likelihood is

$$P(Z_i = 0) + P(Z_i = 1)S^*(T_i = t_i; \lambda | Z_i = 1) = \frac{1}{1 + e^{\beta^T \mathbf{x}_i}} + \frac{e^{\beta^T \mathbf{x}_i}}{1 + e^{\beta^T \mathbf{x}_i}} e^{-\lambda t_i}, \quad t_i \geq 0.$$

Type I censoring was assumed in Farewell (1977). The author also assessed the efficiency of the logistic-exponential mixture model proposed in his paper relative to a regular logistic model, which assumes $Z = 0$ if the patient does not have a relapse until T^0 . He defined T^0 as the end point of the observation time. He found that the efficiency of the mixture model increases as the follow-up time T^0 increases and as $P(Z = 1)$ decreases. When the true underlying distribution is a mixture model, the estimate of the cure rate from the regular logistic model is biased.

Farewell (1982) proposed two more general models for the time to incidence T . He assumed that T follows a Weibull distribution in his first model. The probability density function and the survival function of T given that the patient is

noncured are written as

$$f^*(T = t; \lambda, \zeta \mid Z = 1, \mathbf{x}) = \zeta \lambda (\lambda t)^{\zeta-1} \exp\{-(\lambda t)^\zeta\}, \quad t \geq 0, \quad \text{and} \quad (1.4)$$

$$S^*(T = t \mid Z = 1, \mathbf{x}) = \exp\{-(\lambda t)^\zeta\}, \quad t \geq 0, \quad (1.5)$$

where ζ is an unknown parameter, $\lambda = \exp(-\boldsymbol{\gamma}^T \mathbf{x})$ is a function of covariates, and $\boldsymbol{\gamma}$ is a vector of unknown parameters associated with covariates $\mathbf{x}$. The likelihood function is obtained by plugging in $P(Z = 1)$ defined in Equation (1.3), the probability density function defined in Equation (1.4), and the survival function defined in Equation (1.5) to the likelihood function in Equation (1.2).

Farewell (1982) proposed the second model for grouped data on survival time. Assume there are k failure time groups $(t_i, t_{i+1}]$, $i = 1, 2, \dots, k$, the conditional probability for $T = t_i$ given that $T \geq t_i$ is

$$P(T = t_i \mid Z = 1, T \geq t_i, \mathbf{x}) = \frac{e^{\alpha_i + \boldsymbol{\theta}^T \mathbf{x}}}{1 + e^{\alpha_i + \boldsymbol{\theta}^T \mathbf{x}}}, \quad (1.6)$$

where α_i ($i = 1, 2, \dots, k$) are group-specific parameters. The vector $\boldsymbol{\theta}$ is associated with covariates $\mathbf{x}$ and is common to all time groups. By combining Equations (1.2), (1.3), and (1.6), the likelihood function is obtained.

Farewell (1982) commented that the mixture model is appealing from both biological and statistical points of view. It has been used frequently in cancer studies, especially in the study of breast cancer. Strong evidence of the existence of individuals coming from different populations, however, is required before using the mixture model. Farewell also pointed out that the likelihood function might be flat for some data sets, which would cause possible nonidentifiability of the parameters.

Farewell (1986) discussed some problems related to the mixture model by applying the model to two data sets. The first data set was from a breast cancer trial with 139 patients. He applied a logistic-Weibull mixture model to the data. Then he computed the profile likelihood function for β_0 , the intercept parameter

in the logistic portion of the mixture model. The profile likelihood function for β_0 was obtained by replacing the other parameters in the model by their maximum likelihood estimates (MLE). He found that the profile likelihood function was in a strongly nonquadratic shape, indicating that significant nonidentifiability existed with respect to the estimation of β_0 . Therefore, this logistic-Weibull mixture model presented an unreliable estimate of the cure proportion. Farewell also applied a multinomial-weibull mixture model to a data set from Risch (1983) on the time to the onset of a major affective illness in siblings of affected individuals. Risch's data were grouped into six time intervals. The onset probabilities were modelled by a multinomial distribution. The mixture model fitted data well. Therefore, Farewell concluded that one should be cautious when using mixture models. The cure proportion might be very imprecise in some cases and strong evidence for the existence of two populations is required to make a reliable inference.

A significant amount of literature discussed the mixture model. The common parametric distributions for the time to incidence for the noncured population can take many forms. For example, they can be exponential (Berkson & Gage 1952, Bithell & Upton 1977, Farewell 1977, Goldman 1984); lognormal (Boag 1949); Weibull (Farewell 1982); Gompertz (Gordon 1990); Generalized Gamma (Yamaguchi 1992); and Generalized F-distribution (Peng et al. 1998). The distribution for the time to incidence can also take nonparametric or semiparametric forms, such as Kaplan-Meier (KM) estimates (Laska & Meisner 1992, Taylor 1995) and the proportional hazards model (Kuk & Chen 1992). The probability $P(Z = 1)$ can be treated as a parameter in a Bernoulli trial. It can also be written as a function of a vector of covariates $\mathbf{x}$ in a logistic regression model, such as the model defined in Equation (1.3).

To conclude the previous discussion, it is generally plausible to use the mixture model when there exists strong scientific evidence of the mixture of

different populations. However, as both Farewell (1977, 1982, & 1986) and Cantor and Shuster (1992) pointed out, the likelihood function is sometimes nonquadratic or flat, which could produce problems for both computation and inference. In addition, parameter estimates could fall out of the parameter space when the data do not support a nonzero surviving fraction. The asymptotic properties of the MLE from a logistic-parametric mixture model were developed systematically by Maller and Zhou (1996). The asymptotic properties related to the model parameter estimates in both logistic-semiparametric and logistic-nonparametric models have not been explored in the literature.

1.1.2 Binary Regression Model

Jung (1996) proposed an interesting method to estimate the cure rate, which is defined to be the probability that the time to incidence T is greater than a fixed time τ . In his paper, the standard procedure for inference on parameters in a binary regression model is modified to accommodate censored observations. Let $\phi(\pi) = \beta^T \mathbf{x}$, where ϕ is a link function which is monotonic and differentiable on $[0, 1]$. Also let $\pi^* = \phi^{-1}$, the inverse function of ϕ . Let $\bar{\pi}^* = 1 - \pi^*$ and $\pi_\phi^*(\phi) = \partial\pi^*(\phi)/\partial\phi$. The MLE of β in a standard binary regression model without censoring is obtained by solving

$$n^{-1/2} \sum_{i=1}^n \{I(t_i \geq \tau) - \pi^*(\beta^T \mathbf{x}_i)\} \frac{\pi_\phi^*(\beta^T \mathbf{x}_i)}{\pi^*(\beta^T \mathbf{x}_i) \bar{\pi}^*(\beta^T \mathbf{x}_i)} \mathbf{x}_i = 0.$$

In the case with censoring, denote $T_i^* = \min(T_i, U_i)$. Assume independent censoring, i.e., T_i and U_i are independent given $\mathbf{x}_i$ for $i = 1, 2, \dots, n$. Also assume that $U_1, U_2, \dots, U_n$ are independently, identically distributed (iid) with a cumulative distribution function $1 - G$. Let $\hat{G}$ be the KM estimate of G . Then the MLE $\hat{\beta}_{MLE}$ of β can be obtained by solving

$$n^{-1/2} \sum_{i=1}^n \left\{ \frac{I(t_i^* \geq \tau)}{\hat{G}(\tau)} - \pi^*(\beta^T \mathbf{x}_i) \right\} \frac{\pi_\phi^*(\beta^T \mathbf{x}_i)}{\pi^*(\beta^T \mathbf{x}_i) \bar{\pi}^*(\beta^T \mathbf{x}_i)} \mathbf{x}_i = 0.$$

Numerically $\hat{\beta}_{MLE}$ can be obtained by applying the Newton-Raphson method. Under mild conditions, it is proved that $\hat{\beta}_{MLE}$ is a unique and a consistent estimate of β and $n^{1/2}(\hat{\beta}_{MLE} - \beta)$ follows an asymptotic normal distribution with a zero mean. The problem with this method is that τ is very difficult to define. The result is not convincing if τ is too small. On the other hand, it is not cost effective for τ to be too long.

Farewell (1977) compared the efficiency of a logistic-exponential model to a binary logistic regression model that treats any individual with $t_i^* \geq \tau$ as a long-term survivor. He also pointed out that the estimate of the cure rate would be biased using a logistic model when the underlying true distribution is a mixture distribution. The logistic model discussed by Farewell (1977) did not consider censoring. However, he emphasized that a similar conclusion can be reached by expanding the model to include censored individuals. Farewell also noted that it is especially unreasonable for the logistic model to ignore the possible censoring times when T^* exceeds the largest observing time τ .

1.1.3 Nonparametric Method

Nonparametric tests are widely applied for both comparing survival curves with the presence of cure and comparing cure proportions between treatment arms. When the main interest is to compare survival curves of several populations with the presence of cure, some of the nonparametric tests are considered standard, such as the log-rank test and the generalized Wilcoxon test. Halpern and Brown (1987) studied the power of these two tests by simulations when the mixture cure rate model by Goldman (1984) is appropriate. They compared the power of these two tests for detecting the difference between two treatment arms for selected values of cure rates, instant failure rates, and follow-up time. Each of the two treatment arms had 50 participating individuals. They used 200 replications for each set of simulation. The simulation results are somewhat inconclusive. The log-rank test is

more powerful than the Wilcoxon test in general. However, the log-rank test does not show any advantage over the Wilcoxon test in the following situations: 1) cure rates are high and are at similar levels in both arms; 2) there is a large difference in the time to the end point for both arms; and 3) the follow-up is long relative to the time to failure.

In the comparison of cure proportions, Sposto et al. (1992) studied several tests based on simulated samples. These tests are the likelihood ratio, the Wald, the score tests based on two different mixture models, and a test based on the differences of the KM estimates of the cure rates. The two mixture models are logistic-exponential and logistic-Weibull models proposed by Farewell (1977 & 1982). The time point in the KM estimate of the cure rate is defined as the last observed failure time. Simulated samples under assumed mixture models are generated in order to check the results in an ideal case. In addition, samples under some different models are generated in order to study the robustness of the tests to model misspecification. The authors also generated samples based on various levels of cure proportions. The test statistics calculated from the mixture models sometimes can not be obtained because of the nonconvergence during the maximization procedure. The authors pointed out two reasons for the nonconvergence: 1) the maximization is obtained on the boundary; and 2) the algorithm terminates abnormally with a jump to a nonnegative definite region. Test statistics with convergence problem are not presented in the article. In conclusion, the authors noted that the KM estimate of the cure rate is a reasonable alternative to the parametric estimate in general. When the underlying model includes unequal failure rates between treatment arms, the proportion test based on KM estimates is more powerful than the test based on the complicated parametric model. However, the KM estimate introduces more bias than the parametric

estimate of the cure rate when the sample is small, when failure rates are equal, and when not all failures are observed.

Gray and Tsiatis (1989) proposed a linear rank test to compare survival curves between two treatment arms over the entire time span, but their test focused power against the later difference on survival curves. This linear rank test is evaluated on a mixture distribution of a proportion representing the cure and a failure distribution with an unspecified form. They assumed that failure distribution functions for both treatment arms are constant over time. They specified a weight function in a two-sample non-parametric censored test proposed by Tarone and Ware (1977) in order to obtain their own test. Note that the log-rank test also can be derived from Tarone and Ware's test with a different weight function. However, the difference between Gray and Tsiatis' test and the log-rank test is that the former put more weight towards the later difference and the latter put equal weight throughout the entire observation period.

Gary and Tsiatis (1989) calculated relative efficiencies under equal censoring for their test, the log-rank test, and the proportion test based on KM estimates of cure rates. When there is no censoring, their test gains efficiency over the log-rank test for a small cure rate. However, the efficiency gain is less for a cure rate greater than 50%. These results can be explained by observing that the weight functions become closer for their test and for the log-rank test as the cure rate increases. In addition, under the noncensoring assumption, their test is asymptotically equivalent to the proportion test. When there is censoring and when the censoring proportion increases, their test is increasingly more powerful than the proportion test. But note that their test is not a pure test of the equality of the cure rates. Therefore, it has power against alternatives with unequal cure rates and alternatives with equal cure rates and unequal survival curves.

In this section, two types of statistical tests are discussed: 1) comparing survival curves under the assumption of the existence of cure rates; and 2) comparing cure rates. Some of the nonparametric tests such as the log-rank test is quite standard in oncology for comparing survival curves. When the problem narrows down to the comparison of cure rates, the proportion test based on KM estimates are quite popular due to the computation difficulties related to the test based on the mixture model structure. However, there are problems associated with KM estimates for cure rates, as pointed out by Cantor and Shuster (1992). The standard error of the KM estimate increases in the right-hand extremes because of the small sample size at the end of the study. The KM estimate sometimes can produce counterintuitive results. For example, the time-to-relapse curve plateau could be higher than the time-to-death curve plateau because of the large downward steps at the end of the study. Therefore, one should be cautious relying on the KM estimate of the cure rate. As an alternative to the KM estimate of the cure rate, Cantor and Shuster proposed a parametric model with a bounded cumulative hazard rate to accommodate the cure proportion.

1.1.4 Model with Bounded Cumulative Hazard Rate

Parametric Model

Cantor and Shuster (1992) proposed a Gompertz model to incorporate long-term survivors. The survival function is written as:

$$S(t) = \exp[\beta^{-1}\alpha(1 - e^{\beta t})], \alpha > 0, \quad (1.7)$$

where α is the initial hazard. When $\beta > 0$, the survival function $S(t) \rightarrow 0$ as $t \rightarrow \infty$; and when $\beta < 0$, the survival function $S(t) \rightarrow \exp(\alpha/\beta)$ as $t \rightarrow \infty$. The cure rate is presented as $\exp(\alpha/\beta)$. The authors then maximized the likelihood function based on the Gompertz model using the Newton-Raphson method. Both $\text{Var}(\hat{\alpha})$ and $\text{Var}(\hat{\beta})$ are obtained from the observed information matrix

and the variance of the cure rate $\hat{\alpha}/\hat{\beta}$ is obtained using the Delta method. In the article, they also compared the Gompertz model with the mixture model. When there is a cure proportion, both models produce similar curves and cure rate estimates. However, the Gompertz model is preferred due to the divergence problem associated with the mixture model. The authors also noted the instability of the KM estimate toward the end. Some other nice properties of the proposed Gompertz model are also provided in the article. For example, the cure probability given the patient has survived through time t_0 is $\exp(\frac{\alpha}{\beta}e^{\beta t_0})$; the log-likelihood function is strictly concave so one can easily estimate the parameters and the cure rate; and finally, $S(t)^\Delta$ is also a Gompertz survival function when $S(t)$ is a Gompertz survival function.

Gieser et al. (1998) incorporated covariates information into the model proposed by Cantor & Shuster (1992). The treatment specific survival function for an individual is written as:

$$S(t; i, \mathbf{x}) = \{S_i(t)\}^{e^{\gamma^T \mathbf{x}}} = \exp\{e^{\gamma^T \mathbf{x} + \alpha_i^*} \beta_i^{-1} (1 - e^{\beta_i t})\}, \quad (1.8)$$

where α^* is defined as $\log \alpha$ for α in Equation (1.7), i is the treatment index, $\mathbf{x}$ is a vector of covariates, and γ is a vector of parameters associated with $\mathbf{x}$. Note that $S(t; i, \mathbf{x})$ in Equation (1.8) is also a Gompertz survival function as noted by Cantor and Shuster (1992). The cure rate in this case is

$$\pi(i, \mathbf{x}) = S(\infty; i, \mathbf{x}) = \exp\{\beta_i^{-1} e^{\gamma^T \mathbf{x} + \alpha_i^*}\}.$$

The obtained log-likelihood function is concave with respect to parameters. The MLEs of the parameters are efficient and asymptotic normal when the assumed model is correct and when the sample size is large. The authors compared the treatment effects on cure rates for any given covariates vector $\mathbf{x}$. The ratio of the

log of the cure rates for two different treatment arms is expressed as

$$\theta_{ii'} = \frac{\log \pi(i, \mathbf{x})}{\log \pi(i', \mathbf{x})} = \frac{\beta_{i'} e^{\alpha_i}}{\beta_i e^{\alpha_{i'}}}.$$

The null hypothesis on testing the equality of cure rates is written as

$$H_0 : \log \theta_{ii'} = 0$$

and the test statistic is expressed as

$$Z = [\hat{\alpha}_i + \log(-\hat{\beta}_{i'}) - \log(-\hat{\beta}_i) - \hat{\alpha}_{i'}] / SE,$$

where the standard error SE in the denominator is obtained by the Delta method.

Gieser et al. (1998) applied their model to 763 patients from a multicenter prospective study performed in 1981 by the Pediatric Oncology Group. This study analyzed the standard risk of acute lymphocytic leukemia in children. They compared two treatment effects on the cure rate. Children's age and sex are considered as covariates. The Newton-Raphson algorithm is used to yield MLEs of the parameters. The algorithm converges after nine iterations. Based on this study, they compared the estimates and the variances of the cure rates based on the Gompertz model, the logistic-exponential mixture model, and the KM method. They found that the Gompertz model produces similar estimate of the cure rate to the KM estimate but with much smaller variance. They also provided a generalized model in which the treatment indicator is included as one of the covariates. One important advantage of this generalized model is the log-linear structure of the parameters, since the log-linear structure implies the concavity of the log-likelihood function and the existence, the uniqueness, and the asymptotic normality of the parameter estimates.

Semiparametric Model

Chen et al. (1999) proposed a semiparametric model for the survival data with a cure fraction. They assumed that there are two variables: a) the number of carcinogenic cells N^* in the body of a patient; and b) the incubation time W_i for the i^{th} carcinogenic cell ($i = 1, 2, \dots, N^*$). The number of carcinogenic cells $N^* = 0$ implies that there is no such cell in the patient's body. Assume that N^* follows a Poisson distribution with mean θ and assume that W_i for $i = 1, 2, \dots, N^*$ are iid with distribution function $F(w)$. Variables W_i and N^* are independent. Therefore, the time to incidence is $T = \min(W_i, 0 \leq i \leq N^*)$. Note that $P(W_0 = \infty) = 1$. Then the survival function for T is presented as

$$\begin{aligned}
 S(t) &= P(\text{No incidence by time } t) \\
 &= P(N^* = 0) + P(T \geq t, N^* > 0) \\
 &= P(N^* = 0) + P(W_1 \geq t, W_2 \geq t, \dots, W_{N^*} \geq t, N^* > 0) \\
 &= e^{-\theta} + \sum_{i=1}^{\infty} [1 - F(t)]^i \frac{e^{-\theta}}{i!} \theta^i \\
 &= e^{-\theta} + e^{-\theta} [e^{\theta(1-F(t))} - 1] \\
 &= e^{-\theta F(t)}.
 \end{aligned} \tag{1.9}$$

Note that $S(t)$ is an improper survival function because $S(t) \rightarrow e^{-\theta} > 0$ as $t \rightarrow \infty$. The survival fraction $e^{-\theta}$ is indeed the probability that there exists no carcinogenic cell in a patient's body. The corresponding probability density function and the hazard function are

$$\begin{aligned}
 f(t) &= \theta f(t) \exp(-\theta F(t)), \\
 \lambda(t) &= \theta f(t).
 \end{aligned}$$

Let $S^*(t)$ denote the survival function for the noncured proportion of patients, then

$$S^*(t) = P(T \geq t \mid N^* \geq 1) = \frac{e^{-\theta F(t)} - e^{-\theta}}{1 - e^{-\theta}}. \tag{1.10}$$

The corresponding probability density function is

$$f^*(t) = -\frac{d}{dt}S^*(t) = \frac{e^{-\theta F(t)}}{1 - e^{-\theta}}\theta f(t),$$

and the hazard function is

$$\lambda^*(t) = \frac{f^*(t)}{S^*(t)} = \frac{e^{-\theta F(t)}}{e^{-\theta F(t)} - e^{-\theta}}\theta f(t) = \frac{1}{P(T < \infty | T > t)}\lambda(t).$$

The likelihood function is derived as follows

$$\begin{aligned} L &= \prod_{i=1}^n f(t_i^*)^{\delta_i} S(t_i^*)^{1-\delta_i} \\ &= \prod_{i=1}^n \lambda(t_i^*)^{\delta_i} S(t_i^*) \\ &= \prod_{i=1}^n [\theta f(t_i^*)]^{\delta_i} e^{-\theta F(t_i^*)}. \end{aligned} \quad (1.11)$$

The model is semiparametric in the sense that the form of $F(t)$ is unknown.

It has a strong biological meaning by gathering contributions from two characteristics of the tumor growth, the initial numbers of carcinogenic cells and the rate of their growths. Even without any biological meaning, the model still fits any data with a survival fraction. It can always be treated as being generated by an unknown number N^* of latent competing risks. In this model, $S(t)$, $f(t)$, and $\lambda(t)$ are improper, while $S^*(t)$, $f^*(t)$, and $\lambda^*(t)$ are proper survival function, density function, and hazard function, respectively. One advantage of this model over the mixture model is that the overall hazard function $\lambda(t)$ has a proportional hazard structure if the covariates enter the model through θ . In the mixture model, however, the overall hazard function does not have a proportional hazard structure if the covariates enter the model through the cure rate π . The hazard function $h^*(t)$ for the failure proportion is magnified by $\frac{1}{P(T < \infty | T > t)} > 1$ compared to $h(t)$. As t goes to infinity, function $h^*(t)$ converges to $h(t)$. In addition, the authors pointed out the close relationship between the model presented in Equation (1.9)

and the mixture model presented in Equation (1.1) by observing that $S(t)$ defined in Equation (1.9) can also be written as

$$S(t) = e^{-\theta} + S^*(t)(1 - e^{-\theta}).$$

If we let $P(Z_i = 0)$ in Equation (1.1) take the value of $e^{-\theta}$ and let $S^*(t \mid Z_i = 1)$ take the form of $S^*(t)$ defined in Equation (1.10), then $S(t)$ has the same form as in Equation (1.1). Model in Equation (1.1) also corresponds to some model of the form in Equation (1.9) for some θ and $F(t)$. Note that $F(t)$ is assumed to have a Weibull form for convenience in the article. The parameter θ is a function of covariates in the form of $\theta = \exp(\beta^T \mathbf{x})$.

Tsodikov (1998) proposed the same semiparametric model from different angles. As above, let θ be a function of covariates, denoted by $\theta(\mathbf{x}, \beta)$. Assume that n is the number of failure observations and θ_i^d is the parameter of interest for the i^{th} failure observation. Let m_i be the number of censored observations in the time interval (t_i, t_{i+1}) , where t_i is the failure time associated the i^{th} failure observation. Let $t_{i1}, t_{i2}, \dots, t_{im_i}$ be the censoring time for the observations censored within interval (t_i, t_{i+1}) . Let $\theta_{ij}^c (i = 1, 2, \dots, n, j = 1, 2, \dots, m_i)$ be the parameters of interest for the censored observations. Also let Θ_i be the sum of θ s for all individuals on the interval $[t_i, t_{i+1})$ and let

$$\begin{aligned} S_{a,b} &= \sum_{k=a}^b \Theta_k, \\ \tilde{S}_{a,b} &= \sum_{k=a}^b \theta_k^d, \\ P_{a,b} &= \tilde{S}_{a,b} \tilde{S}_{a+1,b} \cdots \tilde{S}_{b,b}, \text{ and} \\ P^{a,b} &= \tilde{S}_{a,b} \tilde{S}_{a,b-1} \cdots \tilde{S}_{a,a}, \end{aligned}$$

for $0 \leq a \leq b \leq n$. Note that $\sum_a^b = 0$ and $\prod_a^b = 1$ if $a > b$. The form of $F(t)$ is completely unspecified.

In order to draw inference on the cure proportion with the presence of an unknown form of $F(t)$, the author applied the partial likelihood, the marginal likelihood, and the profile likelihood function of θ . The partial likelihood function is written in the form

$$L_p = \frac{\prod_{i=1}^n \theta_i^d}{S_{1,n} S_{2,n} \cdots S_{n,n}}.$$

The partial likelihood is valid due to the proportional hazard structure. It provides consistent estimates. However, it can only be used to estimate the ratio of θ and θ_b , the baseline value. It is therefore inconclusive in the estimation of θ_b . When there is no covariate, i.e., when all patients are homogeneous, cure rate can not be estimated.

Both the marginal and the profile likelihood are derived from the full likelihood function which is written as

$$L = \prod_{i=1}^n \theta_i^d f(t_i) e^{-\theta_i^d F(t_i)} Q(t_i) \prod_{j=1}^{m_i} e^{-\theta_{ij}^c F(t_{ij})} \Delta Q(t_{ij}), \quad (1.12)$$

where $Q(t)$ is the survival function for the censoring time. Type I censoring mechanism with a fixed end point t_c is assumed throughout the article. Then Q is a single-step function such that $Q(t_i) = 1$ for $i = 1, 2, \dots, n$, $\Delta Q(t_c) = 1$, and $t_{nj} = t_c$ for $j = 1, 2, \dots, m_n$. The marginal likelihood function is derived as

$$L_m = \frac{\prod_{i=1}^n \theta_i^d}{S_{1,n} S_{2,n} \cdots S_{n,n}} e^{-(\Theta_i - \theta_i^d) F(t_c)} \sum_{i=0}^n (-1)^i \exp\{-\tilde{S}_{n-i+1,n} F(t_c)\} \frac{P_{1,n}}{P_{1,n-i} P^{n-i+1,n}}.$$

The profile likelihood function is derived from Equation (1.12) by maximizing it with respect to F . Nonparametric maximum likelihood estimator (NPMLE) is used to estimate F .

Both Chen et al. (1999) and Tsodikov (1998) proposed the same semiparametric model, while using different estimation techniques. The former assumed independent and noninformative censoring as well as Weibull distribution for $F(t)$ in the model. Therefore, their model is a parametric model although $F(t)$ can take

a nonparametric form. The latter assumed a Type I censoring mechanism, which is restrictive.

Both the parametric and the semiparametric models with a bounded cumulative hazard rate are appealing because of the simplicity and intuitive meaning of these models. However, there are limitations for the more flexible semiparametric model. It is difficult in terms of computation. In addition, the asymptotic properties of this model have not been explored in the literature. In the next section, we will reexamine the semiparametric model proposed by Chen et al. (1999) and Tsodikov (1998). In this examination, we make no parametric assumption for $F(t)$. In addition, we do not limit the censoring to Type I mechanism. By using the simulated data, we show the computational barrier raised by this model.

1.2 Computation Difficulty of the Semiparametric Model

We assume random and noninformative censoring. The notations used in Tsodikov (1998)'s paper are used in the section except some changes are made in the definition of $P_{a,b}$ and $P^{a,b}$. We define them as follows:

$$\begin{aligned} P_{a,b} &= S_{a,b} S_{a+1,b} \cdots S_{b,b}, \text{ and} \\ P^{a,b} &= S_{a,b} S_{a,b-1} \cdots S_{a,a}. \end{aligned}$$

The likelihood function takes the following form,

$$L = \prod_{i=1}^n \theta_i^d f(t_i) e^{-F(t_i) \Theta_i}, \quad (1.13)$$

where $f(t_i)$ is the density function corresponding to $F(t_i)$. Define

$$I_i^* = (-1)^i \frac{e^{-S_{n-i+1,n}}}{P^{n-i+1,n} P_{1,n-i}},$$

and

$$J_i^* = S'_{n-i+1,n} + \sum_{k=n-i+1}^n \frac{S'_{n-i+1,k}}{S_{n-i+1,k}} + \sum_{k=1}^{n-i} \frac{S'_{k,n-i}}{S_{k,n-i}}, \quad i = 0, 1, \dots, n$$

Define $I_n = \sum_{i=0}^n I_i^*$, $J_n = \sum_{i=0}^n J_i^*$, and $S'_{j,k} = \partial S_{j,k} / \partial \beta$ for any $j, k = 1, 2, \dots, n$.

Let L_m denote the marginal likelihood function. As derived in Appendix A,

$$L_m = \prod_{i=1}^n \theta_i^d I_n. \quad (1.14)$$

Without loss of generality, assume that there is only one covariate x in the study, such as the treatment indicator. Also assume that θ is a function of the covariate x and is expressed in an exponential term as $\exp(\alpha + \beta x)$. The first and the second derivatives of $S_{i,j}$ with respect to β are denoted by $S'_{i,j}$ and $S''_{i,j}$. They are derived as follows:

$$\begin{aligned} S'_{i,j} &= \frac{\partial S_{i,j}}{\partial \beta} = \sum_{k=1}^j (\theta_k^d x_k + \sum_{l=1}^{m_k} \theta_{kl}^c x_{kl}), \\ S''_{i,j} &= \frac{\partial^2 S_{i,j}}{\partial \beta^2} = \sum_{k=1}^j (\theta_k^d x_k^2 + \sum_{l=1}^{m_k} \theta_{kl}^c x_{kl}^2). \end{aligned}$$

The score function for α is obtained as

$$U(\alpha) = - \sum_{i=0}^n \frac{I_i^*}{I_n} S_{n-i+1,n},$$

and the score function for β is written as

$$U(\beta) = \sum_{i=1}^n x_i - \sum_{i=0}^n \frac{I_i^*}{I_n} J_i^*.$$

The Hessian matrix is calculated and written as

$$\begin{bmatrix} \partial^2 \log(L_m) / \partial \alpha^2 & \partial^2 \log(L_m) / \partial \alpha \partial \beta \\ \partial^2 \log(L_m) / \partial \alpha \partial \beta & \partial^2 \log(L_m) / \partial \beta^2 \end{bmatrix},$$

where

$$\partial^2 \log(L_m) / \partial \alpha^2 = - \sum_{i=0}^n S_{n-i+1,n} \frac{I_i^*}{I_n} (-S_{n-i+1,n} + 1) + \frac{1}{I_n^2} \left(\sum_{i=0}^n I_i^* S_{n-i+1,n} \right)^2,$$

$$\partial^2 \log(L_m)/\partial\alpha\partial\beta = \sum_{i=0}^n \frac{I_i^*}{I_n} (J_i^* S_{n-i+1,n} - \sum_{j=0}^n \frac{I_j^*}{I_n} J_j^* S_{n-i+1,n} + S'_{n-i+1,n}),$$

and

$$\partial^2 \log(L_m)/\partial\beta^2 = -\frac{1}{I_n} \sum_{i=0}^n \{-I_i^* J_i^{*2} + I_i^* J_i^{*'}\} + \frac{1}{I_n^2} \left(\sum_{i=0}^n I_i^* J_i^* \right)^2,$$

where $J_i^{*'}$ is the first derivative of J_i^* with respect to β and is expressed as

$$J_i^{*' } = S''_{n-i+1,n} + \sum_{k=n-i+1}^n \frac{S''_{n-i+1,k} S_{n-i+1,k} - (S'_{n-i+1,k})^2}{S_{n-i+1,k}^2} + \sum_{k=1}^{n-i} \frac{S''_{k,n-i} S_{k,n-i} - (S'_{k,n-i})^2}{S_{k,n-i}^2}.$$

Suppose that the covariate x is the treatment indicator, i.e., $x = 1$ if the patient is in the treatment group and 0 if the patient is in the control group. We are now interested in testing the hypothesis:

$$H_0 : \beta = 0. \quad (1.15)$$

The score test (Rao, 1973) is used to test the hypothesis in Equation (1.15).

The score statistic for β under H_0 is denoted as $U(\beta)|_{H_0}$ and the asymptotic variance for the score statistic is $-\partial^2 \log(L_m)/\partial\beta^2$ under the null hypothesis. If α is assumed known, then the standardized test statistic is

$$T = \frac{U(\beta)|_{H_0}}{\sqrt{-\partial^2 \log(L_m)/\partial\beta^2|_{H_0}}}. \quad (1.16)$$

In order to explore the feasibility for computation, we first generate 10000 samples under the assumption that F is an exponential distribution with a mean of 1. Then the survival function is $\exp(-\theta F) = \exp(-(1 - e^{-t})e^{\alpha+\beta x})$. Assume $\alpha = 1.5$ and $\beta = 0$. Each random sample contains 30 observations in the control group and 30 observations in the treatment group. The censoring proportion is around 50%. The standardized score statistic for β is calculated for all random samples generated. The plot of the 10000 standardized score statistics is presented in Figure (1-1). In this bar plot, we observe that the score statistic behaves well and is close to a normal shape given that α is known.

The problem becomes difficult to handle, however, when α is unknown. It is common to use a profile likelihood method to handle the nuisance parameter α when our primary parameter of interest is β . The profile marginal likelihood function for β is obtained from the marginal likelihood function by replacing α by its MLE under H_0 . The test statistic based on the profile marginal likelihood function of β is similar to the one defined in Equation (1.16) given that some adjustment on the denominator of Equation (1.16) is required. Therefore, our primary task is to solve for the MLE of α under H_0 . By using the same parameter setting for Figure (1-1), we generate a sample with 30 patients in the treatment group and 30 in the control group. We show in Figure (1-2) the plot of the log of the marginal likelihood function as a function of α on interval $[-2, 6]$ when β is valued at 0. The marginal likelihood for α under H_0 is poorly behaved. In order to further investigate the behavior of the marginal likelihood function on α and β , we show in Figure (1-3) a three-dimensional plot of the log marginal likelihood function defined based on the same random sample as a function of α on the interval $[-2, 2]$ and β on the interval $[-3, 3]$. The poorly behaved log marginal likelihood as a function of α and β explains the behavior of the log marginal likelihood function in Figure (1-2).

1.3 Piecewise Exponential Model

As we mentioned earlier in the chapter, it is sometimes necessary to assume different hazard functions for the time to failure on different time intervals. Friedman (1981) proposed a piecewise exponential model to handle this problem. Let the time scale be divided into $I(n)$ intervals $(0, T_1], (T_1, T_2], \dots, (T_{I(n)-1}, T_{I(n)}]$. The hazard rate of each individual is assumed to be constant over any given interval. Denote the hazard rate of the j^{th} individual in the i^{th} interval by λ_{ij} . Assume $\lambda_{ij} > 0$ for all (i, j) . Let $l_{ij} = \log \lambda_{ij}$ and let t_{ij} be the survival time of the

j^{th} individual in the i^{th} interval. Define t_{ij} as

$$t_{ij} = \max\{0, \min(T_i - T_{i-1}, t_j - T_{i-1})\}. \quad (1.17)$$

Let

$$I_{ij} = \begin{cases} 1, & \text{when the } j^{th} \text{ individual fails in } [T_{i-1}, T_i) \\ 0, & \text{otherwise.} \end{cases}$$

Then the log-likelihood function is

$$\log L(\mathbf{l}) = \sum_{i,j} l_{ij} I_{ij} - \sum_{i,j} t_{ij} \exp(l_{ij}). \quad (1.18)$$

One of the forms l_{ij} can take is $l_{ij} = \alpha + \sum_k \beta_k x_j^k$.

The author proved the existence and the uniqueness of the MLE of the vector $\mathbf{l}$. The consistency and the asymptotic properties of the MLE were also explored. Karrison (1987) also considered a piecewise exponential model in his study.

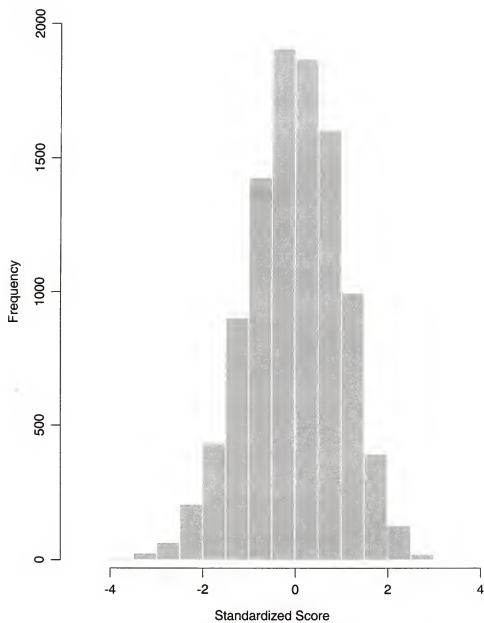


Figure 1-1. Standardized score under H_0 ($\beta = 0$) for 10000 random samples.

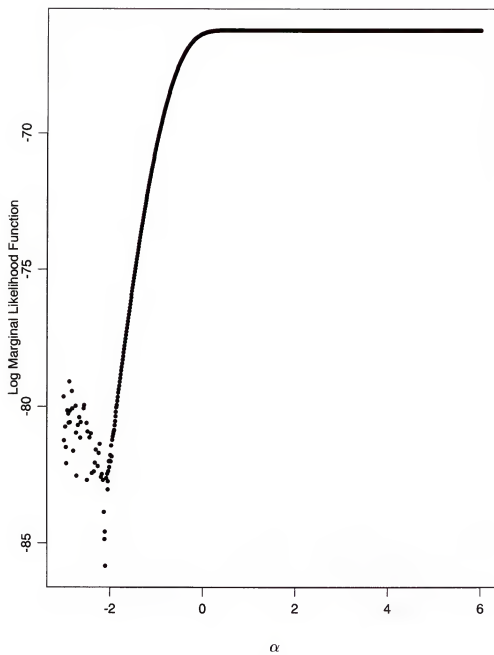


Figure 1-2. Log marginal likelihood function of α when $\beta = 0$.

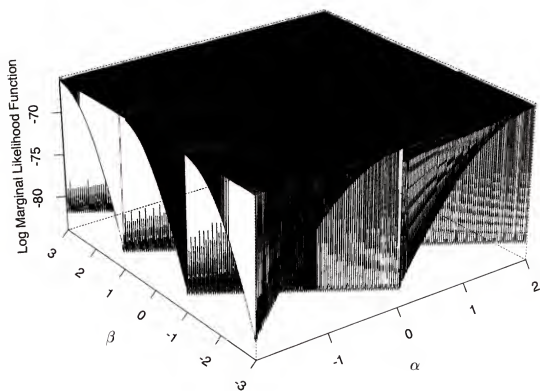


Figure 1-3. Log marginal likelihood function of α and β .

CHAPTER 2

TOPIC OF RESEARCH

The model we propose in this dissertation is a piecewise Gompertz model. It incorporates the cure proportion by assuming a bounded cumulative hazard rate. Moreover, our model gives the flexibility that the survival times for different time intervals may have different hazard functions. As noted by Aitkin et al. (1989) that a piecewise exponential model has the same flexibility as a proportional hazards model under the assumption that each interval is defined by one failure observation in the piecewise model. The choice of using a Gompertz model over an exponential model is made because of the possible plateau of the survival curve from a Gompertz model. We would speculate that under a piecewise Gompertz model, the more intervals we define, the more flexible the model is. Eventually, when there is only one failure observation on each interval, the flexibility of this model would be the same as the semiparametric model we discussed in Chapter 1. Therefore, our model preserves the flexibility of a semiparametric model to some extent. The degree of the flexibility depends on the number of intervals we define for our model. Meanwhile, the computation for our model is much easier to handle than a semiparametric model due to our model's parametric structure. However, when the number of intervals in our model increases, the number of individuals on each interval decreases. We would anticipate large variance for the estimate of the cure rate, which is similar to the situation we discussed earlier for the KM estimate of the cure rate. An appropriate number of intervals needs to be accessed in order to reach maximum flexibility and computational feasibility of the model. In this chapter, we describe our model in detail. In Chapters 3 and 4, asymptotic

properties of the MLEs of parameters in the model are proved. In addition, a goodness-of-fit test is derived which can be applied to determine the appropriate number of time intervals in the model. In Chapter 5, a Monte-Carlo simulation study is carried out to show the advantages of our model. Chapter 6 gives a real example and demonstrates the application of our model to the example.

2.1 Data Structure and Model

Assume there are n patients in the study. Consider the j^{th} patient who is in treatment group g and is associated with a vector of covariates $\mathbf{x}_j$, where $\mathbf{x}_j$ is a p -dimensional vector. Let T_j and U_j be the survival time and the censoring time of the j^{th} patient, respectively. The observed data on the j^{th} patient are $(t_j^*, \delta_j, \mathbf{x}_j)$, where t_j^* is the observation on

$$T_j^* = \min(T_j, U_j),$$

and δ_j is the observation on

$$\Delta_j = I(T_j \leq U_j).$$

Assume there are G treatment groups and the interval $[0, \infty)$ is divided into I intervals $[\tau_{i-1}, \tau_i)$, $i = 1, 2, \dots, I$, where $\tau_0 = 0$ and $\tau_I = \infty$. Let $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)$, where $\boldsymbol{\beta}$ is p -dimensional and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T, \dots, \boldsymbol{\lambda}_G^T)^T$. Note that $\boldsymbol{\lambda}_g^T = (\lambda_{1g}, \lambda_{2g}, \dots, \lambda_{Ig})$. Clearly, $\boldsymbol{\theta}$ is a $(p + IG)$ -dimensional vector. It is assumed that the hazard function for the j^{th} patient in treatment group g has the form

$$e^{\lambda_{ig}t} e^{\boldsymbol{\beta}^T \mathbf{x}_j}, \quad t \in [\tau_{i-1}, \tau_i). \quad (2.1)$$

Define an indicator function $\kappa_g(j)$, where

$$\kappa_g(j) = \begin{cases} 1 & \text{if } j^{\text{th}} \text{ patient is in group } g; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\kappa(j)$ be a vector of $\kappa_g(j)$ for $g = 1, 2, \dots, G$, i.e.,

$$\kappa(j) = (\kappa_1(j), \kappa_2(j), \dots, \kappa_G(j))^T.$$

Let us also define

$$J_i(t) = \begin{cases} 1 & \text{if } t \in [\tau_{i-1}, \tau_i); \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \dots, I$ and

$$\mathbf{J}(t) = (J_1(t), J_2(t), \dots, J_I(t))^T.$$

Let

$$\mathbf{Q}(t, j) = \kappa(j) \otimes \mathbf{J}(t) = \begin{pmatrix} \kappa_1(j)\mathbf{J}(t) \\ \kappa_2(j)\mathbf{J}(t) \\ \vdots \\ \kappa_G(j)\mathbf{J}(t) \end{pmatrix}.$$

Note that $\mathbf{Q}(t, j)$ is a GI -dimensional vector. The hazard function in Equation (2.1) is also written as

$$h(t; \lambda, \kappa(j), \mathbf{J}(t))e^{\beta^T \mathbf{x}_j} = e^{\lambda_{ig}t} e^{\beta^T \mathbf{x}_j}, \text{ if } \kappa_g(j) = 1 \text{ and } J_i(t) = 1.$$

The cumulative hazard function $H_j(t, \theta)$, the survival function $S_j(t, \theta)$, and the density function $f_j(t, \theta)$ can be derived as follows:

$$\begin{aligned} H_j(t; \theta) &= e^{\beta^T \mathbf{x}_j} \int_0^t h(u; \lambda, \kappa(j), \mathbf{J}(u)) du, \\ S_j(t; \theta) &= \exp\{-e^{\beta^T \mathbf{x}_j} \int_0^t h(u; \lambda, \kappa(j), \mathbf{J}(u)) du\}, \\ f_j(t; \theta) &= h(t; \lambda, \kappa(j), \mathbf{J}(t)) e^{\beta^T \mathbf{x}_j} S_j(t; \theta). \end{aligned}$$

The likelihood function contributed by $(t_j, \delta_j, \mathbf{x}_j)$ is expressed as

$$L_{nj} = [h(t_j^*; \lambda, \kappa(j), \mathbf{J}(t_j^*)) e^{\beta^T \mathbf{x}_j}]^{\delta_j} S_j(t_j^*; \theta).$$

The corresponding log-likelihood function is written as

$$\begin{aligned}
 l_{nj} &= \log L_j \\
 &= \delta_j [\log(h(t_j^*; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(t_j^*))) + \boldsymbol{\beta}^T \mathbf{x}_j] - e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du.
 \end{aligned} \tag{2.2}$$

consequently, the overall likelihood function is defined as

$$L_n = \prod_{j=1}^n L_{nj},$$

and the corresponding log-likelihood function is

$$l_n = \sum_{j=1}^n l_{nj}.$$

2.2 Assumptions

Our derivation of asymptotic properties is based on the following three assumptions:

Assumption 1. There is a $B > 0$ such that $\|\mathbf{x}_j\| \leq B$ for $j = 1, 2, \dots, n$.

Assumption 2. There exist $\epsilon > 0$, $\delta > 0$, and an integer n_0 . when $n > n_0$, there exists a subset A of overall individuals with $n_A/n \geq \delta$, where n_A is the size of the subset A such that for any unit vector $\mathbf{a}$,

$$|\mathbf{a}^T \mathbf{x}_j| > \epsilon \quad \text{for } j \in A.$$

Assumption 3. Let n_g be the number of individuals in treatment group g , where $g = 1, 2, \dots, G$. There exists a $\delta > 0$ such that $\min_{1 \leq g \leq G} \frac{n_g}{n} > \delta$ when n is sufficiently large.

Assumption 4. The censoring variables $U_1, U_2, \dots, U_n$ are iid with a survival function $G_U(u)$. Assume that there exists an $\epsilon > 0$ such that

$$G_U(\tau_{I-1}) - G_U(\infty) > \epsilon.$$

The first assumption assumes that all covariates are uniformly bounded. It is reasonable in practice because the majority of the covariates in a cancer study are patients' characteristics, which are bounded uniformly. Note that we are primarily interested in time-independent covariates. The second assumption assumes that at least a set A of patients' covariates are dispersed to some extent. Assumption 3 means that the number of observations in each treatment group must increase proportionally to the sample size. Assumption 4 assumes that not all the censoring observations occur very early on the time span.

2.3 Hypothesis of Interest

We are interested in comparing cure rates among treatment groups. Let $\pi_g(\mathbf{x})$ be the cure rate of patients who are under treatment group g and who are associated with covariates vector $\mathbf{x}$. Considering the case with two treatment groups, the null hypothesis is

$$H_0 : \pi_1(\mathbf{x}) - \pi_2(\mathbf{x}) = 0, \quad (2.3)$$

for any covariate $\mathbf{x}$. The test statistic will be derived in the following section.

2.4 Test of Hypothesis

The cure rate corresponding to patients in treatment group g who are associated with a specific setting of covariates $\mathbf{x}$ is

$$\begin{aligned} \pi_g(\mathbf{x}) &= S(\infty; \theta) = \exp\{-e^{\beta^T \mathbf{x}_j} \int_0^\infty h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du\} \\ &= \exp\{e^{\beta^T \mathbf{x}} \left[\sum_{i=1}^{I-1} \frac{e^{\lambda_{ig} \tau_{i-1}} - e^{\lambda_{ig} \tau_i}}{\lambda_{ig}} + \frac{e^{\lambda_{ig} \tau_{I-1}}}{\lambda_{ig}} \right]\}. \end{aligned} \quad (2.4)$$

The hypothesis defined in Equation (2.3) is equivalent to

$$H_0 : \left\{ \sum_{i=1}^{I-1} \frac{e^{\lambda_{i1} \tau_{i-1}} - e^{\lambda_{i1} \tau_i}}{\lambda_{i1}} + \frac{e^{\lambda_{i1} \tau_{I-1}}}{\lambda_{i1}} \right\} - \left\{ \sum_{i=1}^{I-1} \frac{e^{\lambda_{i2} \tau_{i-1}} - e^{\lambda_{i2} \tau_i}}{\lambda_{i2}} + \frac{e^{\lambda_{i2} \tau_{I-1}}}{\lambda_{i2}} \right\} = 0.$$

Let the MLEs of $\theta = (\lambda^T \beta^T)^T$ be denoted by $\hat{\theta} = (\hat{\lambda}^T \hat{\beta}^T)^T$. The test statistic is derived as

$$T = \left\{ \sum_{i=1}^{I-1} \frac{e^{\widehat{\lambda}_{i1}\tau_{i-1}} - e^{\widehat{\lambda}_{i1}\tau_i}}{\widehat{\lambda}_{i1}} + \frac{e^{\widehat{\lambda}_{I1}\tau_{I-1}}}{\widehat{\lambda}_{I1}} \right\} - \left\{ \sum_{i=1}^{I-1} \frac{e^{\widehat{\lambda}_{i2}\tau_{i-1}} - e^{\widehat{\lambda}_{i2}\tau_i}}{\widehat{\lambda}_{i2}} + \frac{e^{\widehat{\lambda}_{I2}\tau_{I-1}}}{\widehat{\lambda}_{I2}} \right\}. \quad (2.5)$$

The variance of the above test statistic can be derived by the Delta method. The test statistic given in Equation (2.5) is a function of the estimates of λ . We will show in later chapters that $\hat{\theta}$ follows an asymptotic normal with a mean θ and a variance-covariance matrix Σ . The variance estimate of the test statistic can be derived as follows:

$$V(T) = \left(\frac{\partial T}{\partial \lambda} \Big|_{\lambda=\hat{\lambda}} \right)^T \widehat{\Sigma^*} \left(\frac{\partial T}{\partial \lambda} \Big|_{\lambda=\hat{\lambda}} \right),$$

where $\left(\frac{\partial T}{\partial \lambda} \Big|_{\lambda=\hat{\lambda}} \right)$ represents the derivative of the test statistic with respect to λ while evaluated at the MLE of λ . Σ^* is the variance-covariance matrix of $\hat{\lambda}$ and $\widehat{\Sigma^*}$ is the estimate of Σ^* . The vector $\frac{\partial T}{\partial \lambda}$ is derived as follows:

$$\begin{aligned} \frac{\partial T}{\partial \lambda_{i1}} &= \frac{(\tau_{i-1}e^{\lambda_{i1}\tau_{i-1}} - \tau_i e^{\lambda_{i1}\tau_i})\lambda_{i1} - (e^{\lambda_{i1}\tau_{i-1}} - e^{\lambda_{i1}\tau_i})}{\lambda_{i1}^2}, \\ \frac{\partial T}{\partial \lambda_{i2}} &= \frac{(\tau_{i-1}e^{\lambda_{i2}\tau_{i-1}} - \tau_i e^{\lambda_{i2}\tau_i})\lambda_{i2} - (e^{\lambda_{i2}\tau_{i-1}} - e^{\lambda_{i2}\tau_i})}{\lambda_{i2}^2}, \end{aligned}$$

for $i = 1, 2, \dots, I-1$, and

$$\begin{aligned} \frac{\partial T}{\partial \lambda_{I1}} &= \frac{(\lambda_{I1}\tau_{I-1} - 1)e^{\lambda_{I1}\tau_{I-1}}}{\lambda_{I1}^2}, \\ \frac{\partial T}{\partial \lambda_{I2}} &= \frac{(\lambda_{I2}\tau_{I-1} - 1)e^{\lambda_{I2}\tau_{I-1}}}{\lambda_{I2}^2}. \end{aligned}$$

The normalized test statistic is

$$\frac{T}{\sqrt{\widehat{V(T)}}}; \quad (2.6)$$

it follows an asymptotic standard normal distribution.

CHAPTER 3 ESTIMATION OF PARAMETERS

3.1 Likelihood Function and Information Matrices

Based on the log-likelihood function defined in Equation (2.2), the score statistic $S_{nj}(\boldsymbol{\theta})$ for the j^{th} individual can be expressed as

$$S_{nj}(\boldsymbol{\theta}) = \begin{pmatrix} S_{nj}^{(1)}(\boldsymbol{\theta}) \\ S_{nj}^{(2)}(\boldsymbol{\theta}) \end{pmatrix}, \quad (3.1)$$

where

$$S_{nj}^{(1)}(\boldsymbol{\theta}) = \frac{\partial l_{nj}}{\partial \boldsymbol{\beta}} = \delta_j \mathbf{x}_j - \mathbf{x}_j e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \quad (3.2)$$

and

$$S_{nj}^{(2)}(\boldsymbol{\theta}) = \frac{\partial l_{nj}}{\partial \boldsymbol{\lambda}} = \delta_j t_j^* \mathbf{Q}(t_j^*, j) - e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} \mathbf{Q}(u, j) u h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du. \quad (3.3)$$

Note that $S_{nj}^{(1)}(\boldsymbol{\theta})$, $S_{nj}^{(2)}(\boldsymbol{\theta})$, and $S_{nj}(\boldsymbol{\theta})$ are p , IG , and $(p + IG)$ -dimensional vectors, respectively. Furthermore, the matrices formed by second order derivatives of l_{nj} can be derived as

$$\begin{aligned} F_{nj}^{(1,1)}(\boldsymbol{\theta}) &= -\frac{\partial^2 l_{nj}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \mathbf{x}_j \mathbf{x}_j^T e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du, \\ F_{nj}^{(1,2)}(\boldsymbol{\theta}) &= -\frac{\partial^2 l_{nj}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}^T} = \mathbf{x}_j e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} \mathbf{Q}^T(u, j) u h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du, \\ F_{nj}^{(2,1)}(\boldsymbol{\theta}) &= -\frac{\partial^2 l_{nj}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}^T} = e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} \mathbf{Q}(u, j) \mathbf{x}_j^T u h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du, \end{aligned}$$

and

$$F_{nj}^{(2,2)}(\boldsymbol{\theta}) = -\frac{\partial^2 l_{nj}}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T} = e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} \mathbf{Q}(u, j) \mathbf{Q}^T(u, j) u^2 h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du.$$

Then the observed Fisher's information matrix is written as

$$\begin{aligned} F_{nj}(\boldsymbol{\theta}) &= \begin{pmatrix} F_{nj}^{(1,1)}(\boldsymbol{\theta}) & F_{nj}^{(1,2)}(\boldsymbol{\theta}) \\ F_{nj}^{(2,1)}(\boldsymbol{\theta}) & F_{nj}^{(2,2)}(\boldsymbol{\theta}) \end{pmatrix} \\ &= e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} (\mathbf{x}_j^T \ u \mathbf{Q}^T(u, j)) du, \end{aligned} \quad (3.4)$$

which is a $(p + IG) \times (p + IG)$ matrix. The Fisher's information matrix is the expectation of $F_{nj}(\boldsymbol{\theta})$, which is written as

$$D_{nj} = E(F_{nj}(\boldsymbol{\theta})) = e^{\boldsymbol{\beta}^T \mathbf{x}_j} E \left[\int_0^{T_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} (\mathbf{x}_j^T \ u \mathbf{Q}^T(u, j)) du \right]. \quad (3.5)$$

3.2 Concavity of Likelihood Function

The log-likelihood function for n individuals is written as

$$l_n(\boldsymbol{\theta}) = \sum_{j=1}^n \{ \delta_j [\log(h(t_j^*; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(t_j^*))) + \boldsymbol{\beta}^T \mathbf{x}_j] - e^{\boldsymbol{\beta}^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \}. \quad (3.6)$$

The score statistic, the observed Fisher's information matrix, and the Fisher's information matrix for n individuals in the sample are respectively as follows:

$$S_n(\boldsymbol{\theta}) = \sum_{j=1}^n S_{nj}(\boldsymbol{\theta}), \quad (3.7)$$

$$F_n(\boldsymbol{\theta}) = \sum_{j=1}^n F_{nj}(\boldsymbol{\theta}), \quad (3.8)$$

and

$$D_n = \sum_{j=1}^n D_{nj}. \quad (3.9)$$

Let $\mathbf{c}^T = (\mathbf{a}^T \ \mathbf{b}^T)$, where $\mathbf{a}$ is a p -dimensional vector and $\mathbf{b} = (b_{11} \ \cdots \ b_{I1} \ \cdots \ b_{1G} \ \cdots \ b_{IG})^T$ is a IG -dimensional vectors. We will show that $F_n(\boldsymbol{\theta})$ is positive definite and therefore, the log-likelihood function defined in Equation (3.6) is strictly

concave. To show $F_n(\boldsymbol{\theta})$ is positive definite, it suffices to show that $\mathbf{c}^T F_n(\boldsymbol{\theta}) \mathbf{c} > 0$ for any non-null vector $\mathbf{c}$ of dimension $(p + IG)$. In order to prove that $F_n(\boldsymbol{\theta})$ is positive, we assume that there is at least one observation on each interval for each treatment group. Note that

$$\begin{aligned} \mathbf{c}^T F_n(\boldsymbol{\theta}) \mathbf{c} &= \begin{pmatrix} \mathbf{a}^T & \mathbf{b}^T \end{pmatrix} \sum_{j=1}^n F_{nj}(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \\ &= \sum_{j=1}^n e^{\beta^T \mathbf{x}_j} \int_0^{t_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \\ &\geq 0. \end{aligned}$$

This implies that $F_n(\boldsymbol{\theta})$ is semipositive definite. Since $h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) > 0$, $\mathbf{c}^T F_n(\boldsymbol{\theta}) \mathbf{c} = 0$ implies that for any $u \in (0, t_j^*)$, $j = 1, 2, \dots, n$,

$$\mathbf{a}^T \mathbf{x}_j + \mathbf{b}^T \mathbf{Q}(u, j) u = 0. \quad (3.10)$$

For any integer i , $1 \leq i \leq I$ and for any integer g , $1 \leq g \leq G$, there exists an individual j (say) in group g and $t_j^* \in (\tau_{i-1}, \tau_i)$. Then Equation (3.10) implies that

$$\mathbf{a}^T \mathbf{x}_j + u b_{ig} = 0 \text{ for } u \in (\tau_{i-1}, t_j^*).$$

Consequently

$$\mathbf{a}^T \mathbf{x}_j = 0$$

and

$$b_{ig} = 0. \quad (3.11)$$

Since Equation (3.11) holds for any integer i , $1 \leq i \leq I$ and g , $1 \leq g \leq G$, it is then proved that

$$\mathbf{b} = \mathbf{0}.$$

Therefore, Equation (3.10) implies that for any $j = 1, 2, \dots, n$,

$$\mathbf{a}^T \mathbf{x}_j = 0,$$

i.e.,

$$\frac{\mathbf{a}^T}{\|\mathbf{a}\|} \mathbf{x}_j = 0 \text{ for all } j = 1, 2, \dots, n,$$

which is a contradiction to Assumption 2.

Therefore, $F_n(\boldsymbol{\theta})$ is positive definite. The strict concavity of the log-likelihood function as a function of unknown parameters $\boldsymbol{\theta}$ guarantees the uniqueness of the MLEs of the parameters provided that the MLEs exist. Based on the existence of the MLEs and the concavity of the log-likelihood function, the Newton-Raphson algorithm converges.

CHAPTER 4 PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATES

4.1 Introduction

In the previous chapter, we proved the strict concavity of the log-likelihood function. We pointed out that we obtain unique MLEs of parameters if we can prove that these MLEs exist. In this chapter, we will first provide some preliminary results as a preparation for later derivations. Then we will show the existence, the consistency, and the asymptotic normality of the MLEs obtained. The method applied to prove the properties of the MLEs follows a similar method provided by Maller and Zhou (1996, PP211-222). Maller and Zhou proved the asymptotic properties of MLEs in a survival model which assumes that the failure time follows an exponential distribution with covariates information. The exponential model with covariates in their book is a special case of the Gompertz model. By studying the piecewise Gompertz, we move one step further through the incorporation of different hazard functions on different time intervals. In our model, if we allow $\lambda = \mathbf{0}$, the hazard function for the failure time T is $h(t) = e^{\beta x}$. Therefore, the hazard is a constant, which implies that T follows an exponential distribution.

In the next section, we state several lemmas which will be applied to prove the existence, the consistency, and the asymptotic normality of the MLEs. The proofs of the lemmas are given in detail in the Appendices. We prove the existence and the consistency of the MLEs in Section 4.3. The proofs of the asymptotic normality is stated in Section 4.4. In Section 4.5, a goodness-of-fit test which compares models with different numbers of pieces is derived. We also show that the corresponding test statistic follows an asymptotic χ^2 distribution.

4.2 Preliminary Proofs

Lemma 4.1. Let $E(\cdot)$ and $V(\cdot)$ represent the expected value and the variance of a random variable, respectively. Then the following two equations hold:

$$E\{S_n(\boldsymbol{\theta})\} = \mathbf{0} \quad \text{and}$$

$$V\{S_n(\boldsymbol{\theta})\} = D_n.$$

Proof. See Appendix B. □

Lemma 4.2. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0)$ be the true parameter vector, where $\boldsymbol{\lambda}_0 = (\lambda_{11}^{(0)}, \dots, \lambda_{I1}^{(0)}, \dots, \lambda_{1G}^{(0)}, \dots, \lambda_{IG}^{(0)})$. For any positive integer m and for $l = 1, 2$, there exist an $\epsilon > 0$ and a $\Gamma > 0$ such that for any $\boldsymbol{\lambda}'$ satisfies $\|\boldsymbol{\lambda}' - \boldsymbol{\lambda}_0\| < \epsilon$,

$$E_{\boldsymbol{\theta}_0} \left\{ \int_0^{T_j^*} u^m h(u; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\}^l \leq \Gamma,$$

for $j = 1, 2, \dots, n$. Note that Γ is a constant which depends on $\boldsymbol{\theta}_0$ and ϵ .

Proof. See Appendix C. □

Lemma 4.3. Define $\boldsymbol{\theta}_0$ as in Lemma 4.2. For any positive integers l and m , there exists a $\gamma > 0$ such that

$$E_{\boldsymbol{\theta}_0} \left\{ \int_0^{T_j^*} u^m h^l(u; \boldsymbol{\lambda}_0, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\} > \gamma, \quad (4.1)$$

for $j = 1, 2, \dots, n$. Note that γ is a constant depending on $\boldsymbol{\theta}_0$.

Proof. See Appendix D. □

Lemma 4.4. Under Assumptions 2, 3, and 4, there exist a $\gamma > 0$ and an integer $n_0 > 0$ such that when $n \geq n_0$,

$$\frac{\lambda_{\min}(D_n)}{n} \geq \gamma,$$

where $\lambda_{\min}(D_n)$ is the smallest eigenvalue of D_n .

Proof. See Appendix E. □

4.3 Existence and Consistency of the MLE

For each $A > 1$, define an elliptical neighborhood of the true value θ_0 of unknown parameters by

$$N(\theta_0) = \{\theta : (\theta - \theta_0)^T D_n (\theta - \theta_0) \leq A\}. \quad (4.2)$$

As shown in Lemma 4.4, $\lambda_{\min}(D_n) \geq n\gamma$ when $n \geq n_0$. Hereafter we assume that the condition $n \geq n_0$ is satisfied. We observe that for any $\theta \in N(\theta_0)$,

$$\begin{aligned} A \geq (\theta - \theta_0)^T D_n (\theta - \theta_0) &\geq \|\theta - \theta_0\|^2 \lambda_{\min}(D_n) \\ &\geq \|\theta - \theta_0\|^2 n\gamma, \end{aligned}$$

i.e.,

$$\|\theta - \theta_0\| \leq \frac{1}{\sqrt{n}} \sqrt{\frac{A}{\gamma}}. \quad (4.3)$$

Let us consider an arbitrary vector θ which is on the boundary $\partial N(\theta_0)$ of $N(\theta_0)$, then θ is a vector which satisfies the following equation:

$$(\theta - \theta_0)^T D_n (\theta - \theta_0) = A.$$

If we can show that $l_n(\theta) < l_n(\theta_0)$ for all $\theta \in \partial N(\theta_0)$ with probability approaching 1, then the log-likelihood function has a local maximum in $N(\theta_0)$. In addition, the global maximum exists and lies in $N(\theta_0)$ since the log-likelihood function is strictly concave. The consistency of $\widehat{\theta}_n$ is proved by observing that Equation (4.3) holds.

Theorem 4.1. There exists a unique and consistent $\widehat{\theta}_n$ which maximizes $l_n(\theta)$.

Proof. For any $\theta_n \in \partial N_n(\theta_0)$, the Taylor's expansion of the log-likelihood function about θ_0 is written as

$$l_n(\theta_n) - l_n(\theta_0) = (\theta_n - \theta_0)^T S_n(\theta_0) - \frac{1}{2}(\theta_n - \theta_0)^T F_n(\widetilde{\theta}_n)(\theta_n - \theta_0), \quad (4.4)$$

where $\tilde{\theta}_n$ is a random vector on the line segment between θ_n and θ_0 . Hence, the vector $\tilde{\theta}_n \in N(\theta_0)$. Recall that $S_n(\theta_0)$ is the vector of the first-order derivatives of $l_n(\theta)$ evaluated at the true parameter θ_0 and $F_n(\tilde{\theta}_n)$ is the negative matrix of the second derivatives of $l_n(\theta)$ evaluated at $\tilde{\theta}_n$. We have

$$\begin{aligned}
& P\{l_n(\theta_n) - l_n(\theta_0) < 0\} \\
& \geq P\left\{(\theta_n - \theta_0)^T S_n(\theta_0) < \frac{1}{2}(\theta_n - \theta_0)^T F_n(\tilde{\theta}_n)(\theta_n - \theta_0)\right\} \\
& \geq P\left\{(\theta_n - \theta_0)^T S_n(\theta_0) < \frac{1}{4}A, (\theta_n - \theta_0)^T F_n(\tilde{\theta}_n)(\theta_n - \theta_0) > \frac{1}{2}A\right\} \\
& \geq P\left\{(\theta_n - \theta_0)^T F_n(\tilde{\theta}_n)(\theta_n - \theta_0) > \frac{1}{2}A\right\} - P\left\{(\theta_n - \theta_0)^T S_n(\theta_0) \geq \frac{1}{4}A\right\}.
\end{aligned}$$

We proved in Lemma 4.1 that

$$E(S_n(\theta_0)) = 0$$

and

$$V(S_n(\theta_0)) = D_n.$$

Since $(\theta_n - \theta_0)^T S_n(\theta_0)$ is a sum of independent random variables with mean zero, we obtain the following inequality by applying the Chebyshev's inequality:

$$P\{(\theta_n - \theta_0)^T S_n(\theta_0) \geq \frac{A}{4}\} \leq \frac{16(\theta_n - \theta_0)^T D_n(\theta_n - \theta_0)}{A^2} = \frac{16}{A}, \quad (4.5)$$

since $\theta_n \in \partial N(\theta_0)$. The probability in Equation (4.5) can be made arbitrarily small by choosing A sufficiently large. Next we need to show that

$$P\left\{(\theta_n - \theta_0)^T F_n(\tilde{\theta}_n)(\theta_n - \theta_0) > \frac{1}{2}A\right\} \rightarrow 1, \quad (4.6)$$

as $n \rightarrow \infty$.

Let $\mathbf{I}$ be a $(p + IG) \times (p + IG)$ -dimensional identity matrix. If we can show that, uniformly in $\tilde{\theta}_n \in N(\theta_0)$,

$$D_n^{-1/2} F_n(\tilde{\theta}_n) D_n^{-1/2} \xrightarrow{p} \mathbf{I}, \text{ as } n \rightarrow \infty, \quad (4.7)$$

then we obtain the following:

$$\begin{aligned}
& (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)^T F_n(\tilde{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \\
&= (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)^T D_n^{1/2} (D_n^{-1/2} F_n(\tilde{\boldsymbol{\theta}}_n) D_n^{-1/2}) D_n^{1/2} (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \\
&\geq (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)^T D_n (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \lambda_{\min}(D_n^{-1/2} F_n(\tilde{\boldsymbol{\theta}}_n) D_n^{-1/2}) \\
&= A \lambda_{\min}(D_n^{-1/2} F_n(\tilde{\boldsymbol{\theta}}_n) D_n^{-1/2}) \\
&\xrightarrow{P} A,
\end{aligned}$$

where $\boldsymbol{\theta}_n \in \partial N(\boldsymbol{\theta}_0)$ and $\tilde{\boldsymbol{\theta}}_n \in N(\boldsymbol{\theta}_0)$. Therefore, Equation (4.6) holds as $n \rightarrow \infty$. Consequently, if we let both n and A become large,

$$P\{l_n(\boldsymbol{\theta}_n) < l_n(\boldsymbol{\theta}_0)\} \rightarrow 1 \text{ for } \boldsymbol{\theta}_n \in \partial N(\boldsymbol{\theta}_0).$$

Unique MLEs of parameters in the model exist by the strict concavity of the log-likelihood function and are consistent by Equation (4.3). $\square$

We shall now prove Equation (4.7) through the following three lemmas. Note that Equation (4.7) indeed shows that the information matrix D_n is consistently estimated by $F_n(\widehat{\boldsymbol{\beta}}_n)$. Let $\mathbf{c}^T = (\mathbf{a}^T, \mathbf{b}^T)$ be any $(p + IG)$ -dimensional unit vector, where $\mathbf{a}$ and $\mathbf{b}$ are p - and IG -dimensional, respectively.

Lemma 4.5. For any $(p + IG)$ -dimensional unit vector $\mathbf{c}$,

$$\mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}_0) - D_n] D_n^{-1/2} \mathbf{c} \xrightarrow{P} 0, \quad (4.8)$$

where $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\lambda}_0^T)^T$ is the true parameter.

Proof. See Appendix F. $\square$

Lemma 4.6. For any $(p + IG)$ -dimensional unit vector $\mathbf{c}$ and any $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)$,

$$\mathbf{c}^T D_n^{-1/2} (F_n(\boldsymbol{\theta}) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c} \xrightarrow{P} 0.$$

Proof. See Appendix G. $\square$

Lemma 4.7. For any $\boldsymbol{\theta} \in N(\boldsymbol{\theta}_0)$

$$D_n^{-1/2} F_n(\boldsymbol{\theta}) D_n^{-1/2} \xrightarrow{p} \mathbf{I}_{(p+IG) \times (p+IG)} \quad (4.9)$$

Proof. See Appendix H. □

4.4 Asymptotic Normality of the MLE

Lemma 4.8. Let l be an arbitrary positive integer, then $E_{\boldsymbol{\theta}_0} \|S_{nj}(\boldsymbol{\theta})\|^l$ is uniformly bounded for all $j = 1, 2, \dots, n$. Note that $S_{nj}(\boldsymbol{\theta})$ is defined in Equation (3.1).

Proof. See Appendix I. □

Lemma 4.9. For any $(p + IG)$ -dimensional unit vector $\mathbf{c}$,

$$\mathbf{c}^T D_n^{-1/2} S_n(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, 1). \quad (4.10)$$

Proof. Denote

$$Z_n = \mathbf{c}^T D_n^{-1/2} S_n(\boldsymbol{\theta}_0) = \mathbf{c}^T D_n^{-1/2} \sum_{j=1}^n S_{nj}(\boldsymbol{\theta}_0),$$

where $S_{nj}(\boldsymbol{\theta}_0)$, $j = 1, 2, \dots, n$ are independent random variables. By Lemma 4.1

$$E(Z_n) = 0$$

and

$$V(Z_n) = 1.$$

By Liapounov's condition, it suffices to show that

$$\sum_{j=1}^n E_{\boldsymbol{\theta}_0} |\mathbf{c}^T D_n^{-1/2} S_{nj}(\boldsymbol{\theta}_0)|^3 \rightarrow 0 \quad (4.11)$$

as $n \rightarrow \infty$. Note that

$$\begin{aligned} |\mathbf{c}^T D_n^{-1/2} S_{nj}(\boldsymbol{\theta}_0)|^3 &\leq [\lambda_{\max}(D_n^{-1/2})]^3 \|S_{nj}(\boldsymbol{\theta}_0)\|^3 \\ &\leq \left(\frac{1}{\sqrt{n\gamma}}\right)^3 \|S_{nj}(\boldsymbol{\theta}_0)\|^3. \end{aligned}$$

Since $E_{\theta_0} \|S_{nj}(\theta_0)\|^3$ is uniformly bounded for all individuals by Lemma 4.8, Equation (4.11) holds and the proof is accomplished by the Central Limit Theorem. $\square$

Theorem 4.2. The MLE $\widehat{\theta}_n$ are asymptotic normal, i.e.,

$$D_n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}) \text{ as } n \rightarrow \infty.$$

Proof. By Taylor's expansion, we obtain that for some $\tilde{\theta}$ on the line segment between $\widehat{\theta}_n$ and θ_0 and for any unit vector $\mathbf{u}$,

$$\mathbf{u}^T D_n^{-1/2} S_n(\widehat{\theta}_n) = \mathbf{u}^T D_n^{-1/2} S_n(\theta_0) - \mathbf{u}^T D_n^{-1/2} F_n(\tilde{\theta})(\widehat{\theta}_n - \theta_0).$$

Also because $\mathbf{u}^T D_n^{-1/2} S_n(\widehat{\theta}_n) = 0$ since $\widehat{\theta}_n$ is the MLE,

$$\begin{aligned} \mathbf{u}^T D_n^{-1/2} S_n(\theta_0) &= \mathbf{u}^T D_n^{-1/2} F_n(\tilde{\theta})(\widehat{\theta}_n - \theta_0) \\ &= \mathbf{u}^T D_n^{-1/2} (F_n(\tilde{\theta}) - D_n)(\widehat{\theta}_n - \theta_0) + \mathbf{u}^T D_n^{-1/2} D_n(\widehat{\theta}_n - \theta_0) \\ &= \mathbf{u}^T D_n^{-1/2} (F_n(\tilde{\theta}) - D_n) D_n^{-1/2} D_n^{1/2}(\widehat{\theta}_n - \theta_0) + \mathbf{u}^T D_n^{1/2}(\widehat{\theta}_n - \theta_0) \\ &= \mathbf{u}^T [D_n^{-1/2} (F_n(\tilde{\theta}) - D_n) D_n^{-1/2}] D_n^{1/2}(\widehat{\theta}_n - \theta_0) + \mathbf{u}^T D_n^{1/2}(\widehat{\theta}_n - \theta_0). \end{aligned}$$

Since $\widehat{\theta}_n \in N(\theta_0)$, Equation (4.9) holds by Lemma 4.9. Therefore, the sum of the absolute values of all elements of $[D_n^{-1/2} (F_n(\tilde{\theta}) - D_n) D_n^{-1/2}]$ goes to 0 in probability. Hence,

$$[D_n^{-1/2} (F_n(\tilde{\theta}) - D_n) D_n^{-1/2}] = \underline{o_p}(1),$$

where $\underline{o_p}(1)$ is a matrix with all elements converging to 0 as $n \rightarrow \infty$. Then we obtain that

$$\mathbf{u}^T D_n^{-1/2} S_n(\theta_0) = \mathbf{u}^T [\underline{o_p}(1)] D_n^{1/2}(\widehat{\theta}_n - \theta_0) + \mathbf{u}^T D_n^{1/2}(\widehat{\theta}_n - \theta_0).$$

Therefore,

$$\mathbf{u}^T D_n^{1/2}(\widehat{\theta}_n - \theta_0) = \mathbf{u}^T D_n^{-1/2} S_n(\theta_0) + \underline{o_p}(1).$$

Since we have proved the normality of $\mathbf{u}^T D_n^{-1/2} S_n(\boldsymbol{\theta}_0)$ in Lemma 4.7, the normality of $\mathbf{u}^T D_n^{-1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ follows. Hence, Theorem 4.2 is proved. $\square$

4.5 Goodness-of-Fit Test

In practice, one might be interested in knowing how many intervals are in fact necessary in a specific study. Therefore, a goodness-of-fit test can be performed to test how many time intervals are adequate. Assume that we have a saturated model with $I_1 + I_2$ intervals and an unsaturated model with I_1 intervals. We are interested in testing whether there is significant difference between the two models. We assume that the unsaturated model is obtained by combining intervals $I_1 + 1, I_1 + 2, \dots, I_1 + I_2$ in the saturated model with interval I_1 . The null hypothesis can be written as

$$H_0 : \lambda_{(I_1+1)g} = \lambda_{(I_1+2)g} = \dots = \lambda_{(I_1+I_2)g} = \lambda_{I_1g}, \quad \forall g = 1, 2, \dots, G. \quad (4.12)$$

Since $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T \quad \boldsymbol{\lambda}^T)$, the null hypothesis defined above can also be written as

$$H_0 : H\boldsymbol{\theta} = \mathbf{0}, \quad (4.13)$$

where H is a design matrix with full row rank $I_2 G$ and is written as

$$\begin{pmatrix} \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 1 & -1 & 0 & \cdots & 0 & \cdots & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 \\ \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 1 & 0 & -1 & \cdots & 0 & \cdots & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 1 & 0 & 0 & \cdots & -1 & \cdots & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 & \cdots & \mathbf{0}_{I_1-1}^T & 1 & -1 & 0 & \cdots & 0 \\ \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 & \cdots & \mathbf{0}_{I_1-1}^T & 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p^T & \mathbf{0}_{I_1-1}^T & 0 & 0 & 0 & \cdots & 0 & \cdots & \mathbf{0}_{I_1-1}^T & 1 & 0 & 0 & \cdots & -1 \end{pmatrix},$$

where $\mathbf{0}$ s in bold are null vectors. Note that the first p columns of the matrix H put no restriction on β . Columns $(p + 1)$ through $(p + I_1 - 1)$ represent the nonrestrictiveness of λ for treatment group 1 corresponding to the first $I_1 - 1$ intervals. Similar patterns are presented for the rest of the columns. We will now construct a likelihood ratio test for the hypothesis in Equation (4.13). To maximize the log-likelihood function $l_n(\theta)$ under H_0 is equivalent to maximize

$$l_n(\theta) - \theta^T H^T \gamma$$

unrestrictively in $\theta \in R^{[p+(I_1+I_2)G]}$ and $\gamma \in R^{I_2 G}$, where γ is a vector of Lagrange multipliers.

If we let θ_0 be the true parameters, $\dot{\theta}_n$ be the parameters estimated under H_0 , and $\hat{\theta}_n$ be the estimated parameters under the saturated model, then $\dot{\theta}_n^T$ satisfies the following equation:

$$S_n(\dot{\theta}_n) - H^T \dot{\gamma} = \mathbf{0}, \quad (4.14)$$

where $\dot{\gamma}$ is a vector of the estimated γ under H_0 . Also under the null hypothesis, $H\dot{\theta}_n = \mathbf{0}$ and $H\theta_0 = \mathbf{0}$,

$$H(\dot{\theta}_n - \theta_0) = \mathbf{0}. \quad (4.15)$$

Now we need to solve for $\dot{\theta}_n$ from Equations (4.14) and (4.15). By Taylor's expansion,

$$D_n^{-1/2} S_n(\dot{\theta}_n) = D_n^{-1/2} S_n(\theta_0) - D_n^{-1/2} (\dot{\theta}_n - \theta_0) + o_p(1),$$

hence

$$D_n^{1/2} (\dot{\theta}_n - \theta_0) + D_n^{-1/2} S_n(\dot{\theta}_n) = D_n^{-1/2} S_n(\theta_0) + o_p(1).$$

By Equation (4.14), the above equation can be written as

$$D_n^{1/2} (\dot{\theta}_n - \theta_0) + D_n^{-1/2} H^T \dot{\gamma} = D_n^{-1/2} S_n(\theta_0) + o_p(1). \quad (4.16)$$

We can rewrite Equation (4.15) as

$$(HD_n^{-1/2})D_n^{1/2}(\dot{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \mathbf{0}. \quad (4.17)$$

Now express Equations (4.16) and (4.17) in a matrix formula,

$$\begin{pmatrix} I & D_n^{-1/2}H^T \\ HD_n^{-1/2} & \mathbf{0} \end{pmatrix} \begin{pmatrix} D_n^{1/2}(\dot{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ \dot{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} D_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1) \\ \mathbf{0} \end{pmatrix}. \quad (4.18)$$

Since H is of full row rank I_2G and D_n is nonsingular, $HD_n^{-1}H^T$ is of full rank I_2G . Hence, the inverse of matrix $\begin{pmatrix} I & D_n^{-1/2}H^T \\ HD_n^{-1/2} & \mathbf{0} \end{pmatrix}$ exists and can be written as

$$\begin{pmatrix} I & D_n^{-1/2}H^T \\ HD_n^{-1/2} & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} I - P & Q \\ Q^T & R \end{pmatrix},$$

where

$$P = D_n^{-1/2}H^T(HD_n^{-1}H^T)^{-1}HD_n^{-1/2},$$

$$Q = D_n^{-1/2}H^T(HD_n^{-1}H^T)^{-1}, \text{ and}$$

$$R = -(HD_n^{-1}H^T)^{-1}.$$

Therefore, we can solve Equations (4.18) as

$$\begin{pmatrix} D_n^{1/2}(\dot{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ \dot{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} I - P & Q \\ Q^T & R \end{pmatrix} \begin{pmatrix} D_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1) \\ \mathbf{0} \end{pmatrix},$$

hence,

$$D_n^{1/2}(\dot{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = (I - P)D_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1).$$

Similarly for $\hat{\boldsymbol{\theta}}_n$, we have

$$D_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = D_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1),$$

by Taylor's expansion and by observing the fact that $S_n(\hat{\boldsymbol{\theta}}_n) = 0$. Therefore,

$$\begin{aligned} D_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \dot{\boldsymbol{\theta}}_n) &= D_n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - D_n^{1/2}(\dot{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &= PD_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1). \end{aligned} \quad (4.19)$$

Note that $PP = P$ is idempotent with full rank I_2G ; therefore, P can be written as

$$P = A^T \begin{pmatrix} I_{I_2G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} A,$$

where A is an orthogonal matrix. Let $Z_n = AD_n^{-1/2}S_n(\boldsymbol{\theta}_0)$, then

$$Z_n \xrightarrow{d} N(\mathbf{0}, I),$$

by Lemma 4.9. From Equation (4.19), we obtain that

$$\begin{aligned} &(\hat{\boldsymbol{\theta}}_n - \dot{\boldsymbol{\theta}}_n)^T D_n(\hat{\boldsymbol{\theta}}_n - \dot{\boldsymbol{\theta}}_n) \\ &= S^T(\boldsymbol{\theta}_0)D_n^{-1/2}PD_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1) \\ &= S^T(\boldsymbol{\theta}_0)D_n^{-1/2}A^T \begin{pmatrix} I_{I_2G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} AD_n^{-1/2}S_n(\boldsymbol{\theta}_0) + o_p(1) \\ &= Z_n^T \begin{pmatrix} I_{I_2G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Z_n + o_p(1) \\ &\rightarrow Z^T \begin{pmatrix} I_{I_2G} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Z + o_p(1) \\ &\sim \mathcal{X}_{I_2G}^2. \end{aligned} \quad (4.20)$$

The likelihood ratio test statistic for the hypothesis in Equation (4.12) can be written as

$$\Lambda = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L_n(\boldsymbol{\theta})}{\sup_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta})} = \frac{L_n(\dot{\boldsymbol{\theta}}_n)}{L_n(\hat{\boldsymbol{\theta}}_n)},$$

where Θ represents the overall parameter space and Θ_0 the parameter space under H_0 . By Taylor's expansion, we can express $-2 \log(\Lambda)$ as

$$\begin{aligned} -2 \log(\Lambda) &= -2[l_n(\dot{\theta}_n) - l_n(\hat{\theta}_n)] \\ &= -2[(\dot{\theta}_n - \hat{\theta}_n)^T S_n(\hat{\theta}_n)] + (\dot{\theta}_n - \hat{\theta}_n)^T F_n(\tilde{\theta})(\dot{\theta}_n - \hat{\theta}_n), \end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}_n$ and $\dot{\theta}_n$ and therefore $\tilde{\theta} \in N_n(A)$. Note that $S_n(\hat{\theta}_n) = 0$ and $F_n(\tilde{\theta})$ can be replaced by D_n from Lemma 4.7. Hence, $-2 \log(\Lambda)$ is written as

$$-2 \log(\Lambda) = (\dot{\theta}_n - \hat{\theta}_n)^T D_n(\dot{\theta}_n - \hat{\theta}_n) + o_p(1),$$

which converges in distribution to $\mathcal{X}_{I_2 G}^2$ by Equation (4.20).

CHAPTER 5 SIMULATIONS

5.1 Introduction

The results discussed in the previous chapters showed the asymptotic properties of parameter MLEs in the piecewise Gompertz model. These results enable us to carry out the hypothesis test of which the null hypothesis is proposed in Equation (2.3). In this chapter, Monte Carlo simulation studies are performed in order to evaluate the normality approximation of the proposed test statistic in Equation (2.6) and the power of the test under different settings based on the amount of location shift and the sample size. Due to the complexity of the piecewise model, the procedure for generating the random sample is quite complicated and is discussed in section 5.2. Results are given on observed Type I errors and powers for both the piecewise Gompertz model and the exponential-logistic mixture model. All simulations were completed on the Sun Ultra, Enterprise 450 computer using Ox 2.0 (Doornik 1998). Random number generators in the Ox 2.0 environment were used to generate uniformly distributed random numbers.

5.2 Generation of Random Samples

In this section we describe the procedure for generating the random sample when the sample follows a piecewise Gompertz model based on certain parameter settings. In order to generate a random sample of size n with a specified proportion of censoring, we need to generate a sample of failure observations of size n and a sample of censored observations of size n . The sample of failure observations follows the piecewise Gompertz model. The sample of censored observations

is assumed to follow a uniform $(0, M)$ distribution, where M is defined by the censoring proportion. The final sample is obtained by comparing the two samples generated. A subject is considered failed if its observation from the first sample is smaller than the one obtained from the second sample. The subject is censored if its observation from the first sample is greater than the one from the second sample.

5.2.1 Parameter Settings

We consider five different parameter settings in our Monte Carlo simulations. There is no treatment difference in random samples generated based on Settings 1 and 2. Therefore we can compare Type I errors between tests based on random samples generated from these two settings. In parameter Settings 3, 4, and 5, treatment effects exist between groups. Settings 3 to 5 present the smallest, the intermediate, and the largest difference between treatment groups, respectively. The comparison of powers of tests are based on random samples generated from these three parameter settings. All five parameter settings are listed as follows: Setting 1:

- Two treatment groups
- Three intervals: $[0, 3)$, $[3, 7)$, and $[7, \infty)$
- One covariate with value 0 or 1
- Cure proportion varied with covariate values
- Censoring rates: 30%
- $\lambda_{10} = \lambda_{11} = -4.2$
- $\lambda_{20} = \lambda_{21} = -0.52$
- $\lambda_{30} = \lambda_{31} = -0.2$
- $\beta_0 = -0.1$, $\beta_1 = 0.2$

Setting 2:

- Two treatment groups

- Three intervals: $[0, 4)$, $[4, 10)$, and $[10, \infty)$
- One covariate with value 0 or 1
- Cure proportion varied with covariate values
- Censoring rates: 30%
- $\lambda_{10} = \lambda_{11} = -3.5$
- $\lambda_{20} = \lambda_{21} = -0.45$
- $\lambda_{30} = \lambda_{31} = -0.15$
- $\beta_0 = -0.2$, $\beta_1 = 0.15$

Setting 3:

- Two treatment groups
- Three intervals: $[0, 3)$, $[3, 7)$, and $[7, \infty)$
- One covariate with value 0 or 1
- Cure proportion varied with covariate values and treatment groups
- Censoring rates: 30%
- $\lambda_{10} = -3.5$ and $\lambda_{11} = -4.2$
- $\lambda_{20} = -0.45$ and $\lambda_{21} = -0.52$
- $\lambda_{30} = -0.21$ and $\lambda_{31} = -0.2$
- $\beta_0 = -0.1$, $\beta_1 = 0.2$

Setting 4:

- Two treatment groups
- Three intervals: $[0, 4)$, $[4, 10)$, and $[10, \infty)$
- One covariate with value 0 or 1
- Cure proportion varied with covariate values and treatment groups
- Censoring rates: 30%
- $\lambda_{10} = -3.9$ and $\lambda_{11} = -4.2$
- $\lambda_{20} = -0.45$ and $\lambda_{21} = -0.4$
- $\lambda_{30} = -0.165$ and $\lambda_{31} = -0.18$

- $\beta_0 = -0.2$, $\beta_1 = 0.15$

Setting 5:

- Two treatment groups
- Three intervals: $[0, 4)$, $[4, 10)$, and $[10, \infty)$
- One covariate with value 0 or 1
- Cure proportion varied with covariate values and treatment groups
- Censoring rates: 30%
- $\lambda_{10} = -3.5$ and $\lambda_{11} = -4.2$
- $\lambda_{20} = -0.45$ and $\lambda_{21} = -0.4$
- $\lambda_{30} = -0.15$ and $\lambda_{31} = -0.18$
- $\beta_0 = -0.2$, $\beta_1 = 0.15$

The survival curves for the above five parameter settings are shown in Figures (5-1) through (5-5). The survival probabilities are calculated by applying

$$S(t) = \exp\left\{e^{\beta^T \mathbf{x}} \left(\sum_{m=1}^{i-1} \frac{e^{\lambda_{mg}\tau_{m-1}} - e^{\lambda_{mg}\tau_m}}{\lambda_{mg}} + \frac{e^{\lambda_{ig}\tau_{i-1}} - e^{\lambda_{ig}t}}{\lambda_{ig}} \right) \right\}, \quad (5.1)$$

for $t \in [\tau_{i-1}, \tau_i)$. There are two separate survival curves in the first two figures for individuals with covariate values of 0 and 1. For Figures (5-3), (5-4), and (5-5), we draw four survival curves for all four possible combinations of the covariate and the treatment group indicator.

5.2.2 Generation of Censored Observations

From the assumption, the sample of censored observations follows a uniform $(0, M)$ distribution. Therefore, we can obtain a sample with a uniform $(0, 1)$ distribution and then transform them into a sample that follows a uniform $(0, M)$ distribution. The primary task is to obtain an appropriate value for M . Let U and T denote random variables for censored and failure observations, respectively. By taking parameter setting 1 as an example, we illustrate the derivation of the

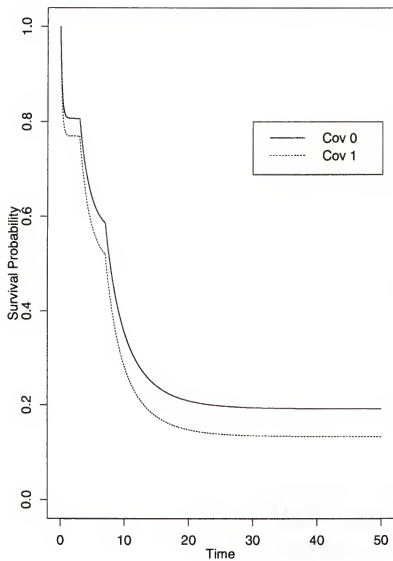


Figure 5-1. Survival curves for patients with parameter setting 1.

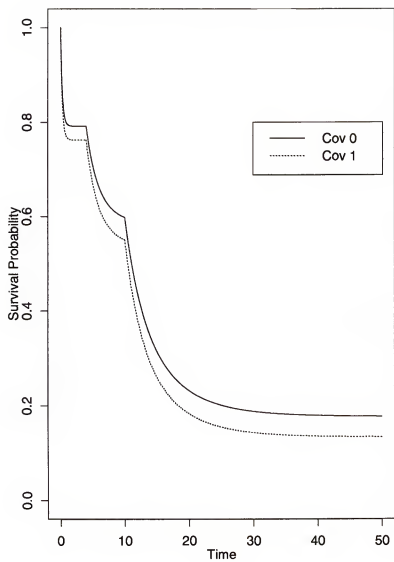


Figure 5-2. Survival curves for patients with parameter setting 2.

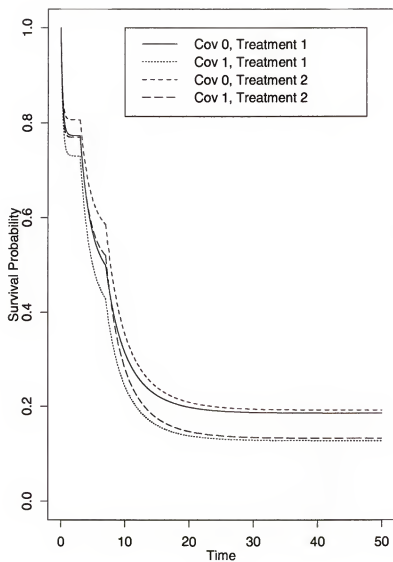


Figure 5-3. Survival curves for patients with parameter setting 3.

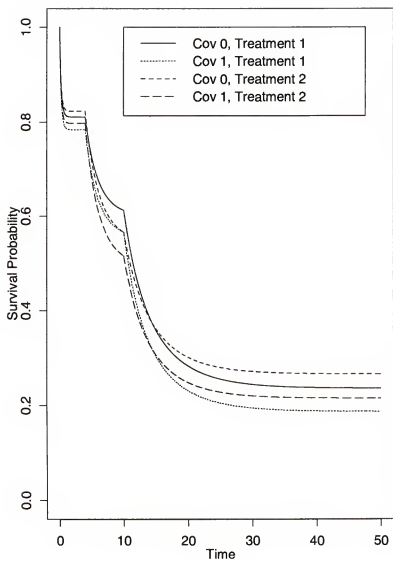


Figure 5-4. Survival curves for patients with parameter setting 4.

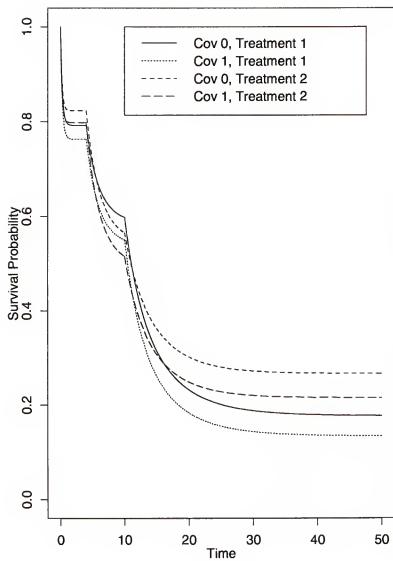


Figure 5-5. Survival curves for patients with parameter setting 5.

censoring proportion $P(U < T)$ as follows.

$$P(U < T) = \frac{1}{M} \int_0^M e^{-\frac{e^{\beta_0 + \beta_1 x}}{\lambda_{1g}}(e^{\lambda_{1g}c} - 1)} dc,$$

for $M \in [0, 3]$;

$$\begin{aligned} P(U < T) &= \frac{1}{M} \int_0^3 \exp\left\{e^{\beta_0 + \beta_1 x} \frac{1 - e^{\lambda_{1g}c}}{\lambda_{1g}}\right\} dc \\ &+ \frac{1}{M} \exp\left\{e^{\beta_0 + \beta_1 x} \frac{1 - e^{\lambda_{1g}3}}{\lambda_{1g}}\right\} \int_3^M \exp\left\{e^{\beta_0 + \beta_1 x} \frac{e^{\lambda_{2g}3} - e^{\lambda_{2g}c}}{\lambda_{2g}}\right\} dc, \end{aligned}$$

for $M \in [3, 7]$; and for $M \in [7, \infty)$,

$$\begin{aligned} P(U < T) &= \frac{1}{M} \int_0^3 \exp\left\{e^{\beta_0 + \beta_1 x} \frac{1 - e^{\lambda_{1g}c}}{\lambda_{1g}}\right\} dc \\ &+ \frac{1}{M} \exp\left\{e^{\beta_0 + \beta_1 x} \frac{1 - e^{\lambda_{1g}3}}{\lambda_{1g}}\right\} \int_3^7 \exp\left\{e^{\beta_0 + \beta_1 x} \frac{e^{\lambda_{2g}3} - e^{\lambda_{2g}c}}{\lambda_{2g}}\right\} dc \\ &+ \frac{1}{M} \exp\left\{e^{\beta_0 + \beta_1 x} \left[\frac{1 - e^{\lambda_{1g}3}}{\lambda_{1g}} + \frac{e^{\lambda_{2g}3} - e^{\lambda_{2g}7}}{\lambda_{2g}}\right]\right\} \int_7^M \exp\left\{e^{\beta_0 + \beta_1 x} \frac{e^{\lambda_{3g}7} - e^{\lambda_{3g}c}}{\lambda_{3g}}\right\} dc. \end{aligned}$$

Calculation of integrals in the above equations are carried out approximately using the Simpson's Formula. The detail of the Simpson's Formula is provided in Appendix J. Therefore, we search for an appropriate value for M such that $P(U < T)$ reaches the required censoring proportion.

5.2.3 Generation of Failure Observations

In order to generate a sample of failure observations, we need to generate a sample of uniform $(0, 1)$ observations. Then we can treat the value from the uniform $(0, 1)$ as the survival probability for the failure time T . By inverting the survival function defined in Equation (5.1), we obtain the value for the failure time. Note that special attention is required at the inversion because we are dealing with

piecewise distribution. Examples are given for observations with covariate $x = 0$ in the first parameter setting.

From Figure (5-1), we observe that each survival curve consists of three pieces based on the three intervals we consider in the study. Therefore, the inversion function for generating the failure time observation depends on the value of the generated uniform random number U^* . The survival probability when $t = 0$ is 1, the survival probability

$$P(T > 3) = \exp\{e^{-0.1} \frac{1 - e^{-4.2 \times 3}}{-4.2}\} = 0.81,$$

and

$$P(T > 7) = \exp\{e^{-0.1} [\frac{1 - e^{-4.2 \times 3}}{-4.2} + \frac{e^{-0.52 \times 3} - e^{-0.52 \times 7}}{-0.52}]\} = 0.59.$$

Therefore if $u^* \in [0.81, 1]$, then

$$T = \frac{1}{\lambda_{1g}} \log[e^{\lambda_{1g}\tau_0} - \lambda_{1g}e^{-\beta_0} \log(u^*)];$$

if $u^* \in [0.59, 0.81)$, then

$$T = \frac{1}{\lambda_{2g}} \log\left\{e^{\lambda_{2g}\tau_1} - \lambda_{2g}\left[e^{-\beta_0} \log(u^*) - \frac{e^{\lambda_{1g}\tau_0} - e^{\lambda_{1g}\tau_1}}{\lambda_{1g}}\right]\right\};$$

and if $u^* \in [0, 0.59)$, then

$$T = \frac{1}{\lambda_{3g}} \log\left\{e^{\lambda_{3g}\tau_2} - \lambda_{3g}\left[e^{-\beta_0} \log(u^*) - \left(\frac{e^{\lambda_{1g}\tau_0} - e^{\lambda_{1g}\tau_1}}{\lambda_{1g}} + \frac{e^{\lambda_{2g}\tau_1} - e^{\lambda_{2g}\tau_2}}{\lambda_{2g}}\right)\right]\right\}.$$

Note that it is necessary that there is at least one failure observation in each interval for each treatment group in order to make inferences on all parameters.

5.3 Type I Error and Power

Type I errors and powers of tests are calculated based on the piecewise Gompertz model and on the exponential-logistic mixture model for all five sets of parameters. The test statistic based on the piecewise Gompertz model

proposed in this dissertation, as we recall, was presented in Equation (2.6). For the exponential-logistic mixture model, we assume that there is one covariate x in the model. The time to failure is assumed to follow an exponential distribution with mean $1/\lambda$, where λ is modelled as a function of the covariate x and the treatment indicator g . The parameter λ is presented as $\exp(\gamma_0 + \gamma_1 x + \gamma_2 g)$. The cure rate $P(Z = 0)$ is defined as $\frac{1}{1 + e^{\beta_0 + \beta_1 x + \beta_2 g}}$. Therefore, for a fixed covariate value, the test of the equality of cure rates between two treatment groups is equivalent to the test of $H_0 : \beta_2 = 0$. Hence, the standardized test statistic for this null hypothesis is

$$TS = \frac{\hat{\beta}_2}{\sqrt{V(\hat{\beta}_2)}}.$$

Type I errors based on the first two parameter settings for both models are shown in Table (5-1). Powers for tests based on parameter settings 3 through 5 for both models are presented in Table (5-2). All tests are carried out for 10000 generated random samples and for three different sample sizes: 400, 1000, and 2000. The numbers of unobtainable test statistics (# NaNs) for the exponential-logistic mixture model are shown in the last column of each table. The unobtainable test statistics occur because the likelihood function for the mixture model is close to flat at the point that MLEs are obtained. When the sample size becomes larger, however, unobtainable or unidentifiable test statistics are eliminated.

We present bar plots of all standardized test statistics obtained for all three sample sizes based on those five parameter settings. These plots help to visualize the advantages of the piecewise Gompertz model in terms of the normal approximation of the test under the null hypothesis and the power of the test under the alternative hypothesis. Figures (5-6), (5-7), and (5-8) are the distributions of the test statistics under parameter setting 1 for both the mixture model and the piecewise Gompertz model. These three figures use sample sizes of 400, 1000, and

Table 5-1. Type I Errors for the Test of Equality of Cure Proportions for the Mixture and the Three-Piece Gompertz Models.

Parameter Setting	Sample Size	Type I Error		# NaNs	
		Mixture	Gompertz	Mixture	Gompertz
1	(200, 200)	0.0305	0.0375	220	
	(500, 500)	0.0171	0.0412	1	
	(1000, 1000)	0.0365	0.0412		
2	(200, 200)	0.0203	0.0373	69	
	(500, 500)	0.0356	0.0360		
	(1000, 1000)	0.0402	0.0408		

Table 5-2. Powers for Tests of Equality of Cure Proportions for the Mixture and the Three-Piece Gompertz Models.

Parameter Setting	Sample Size	Power		# NaNs	
		Mixture	Gompertz	Mixture	Gompertz
3	(200, 200)	0.0260	0.1016	127	
	(500, 500)	0.0328	0.2024		
	(1000, 1000)	0.0679	0.3703		
4	(200, 200)	0.0616	0.1896	5	
	(500, 500)	0.0357	0.4351		
	(1000, 1000)	0.2721	0.7379		
5	(200, 200)	0.2706	0.5572	28	
	(500, 500)	0.7180	0.9331		
	(1000, 1000)	0.9601	0.9981		

2000, respectively. Similarly, Figures (5-9), (5-10), and (5-11) are for parameter setting 2; Figures (5-12), (5-13), and (5-14) are for parameter setting 3; Figures (5-15), (5-16), and (5-17) are for parameter setting 4; and Figures (5-18), (5-19), and (5-20) are for parameter setting 5.

From these tables and plots, we observe that Type I errors obtained from the three-piece Gompertz model are closer to 0.05 than those from the mixture model. Powers presented in Table (5-2) for the three-piece Gompertz model are much larger than ones for the mixture model for all sample sizes. The estimation procedure for obtaining MLEs is also easier to carry out for the three-piece Gompertz model than for the exponential-logistic model. In addition, the algorithm for the three-piece Gompertz model converges much faster.

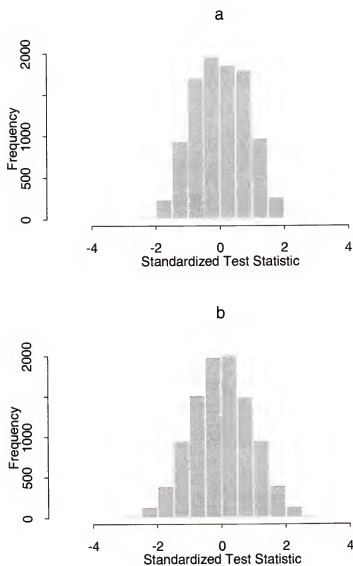


Figure 5-6. Standardized test statistics for parameter setting 1 for 10000 random samples with Size 400. a) Mixture model; b) Three-piece Gompertz model.

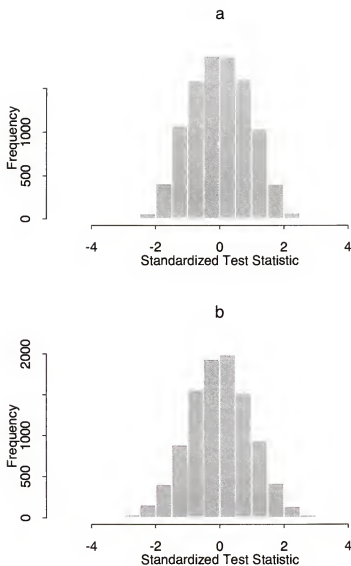


Figure 5-7. Standardized test statistics for parameter setting 1 for 10000 random samples with size 1000. a) Mixture model; b) Three-piece Gompertz model.

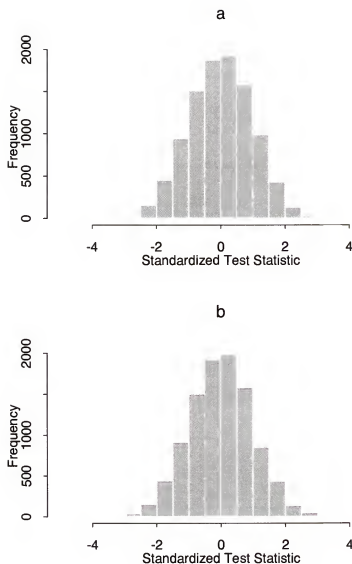


Figure 5-8. Standardized test statistics for parameter setting 1 for 10000 random samples with size 2000. a) Mixture Model; b) Three-piece Gompertz model.

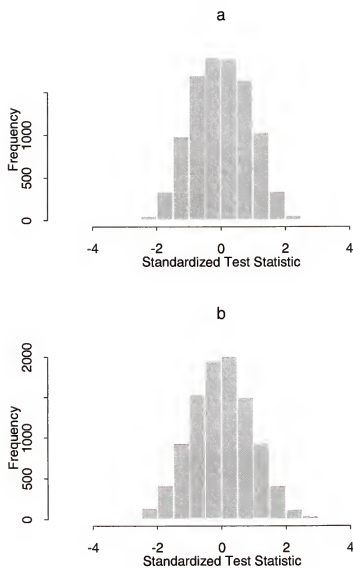


Figure 5-9. Standardized test statistics for parameter setting 2 for 10000 random samples with size 400. a) Mixture model; b) Three-piece Gompertz model.

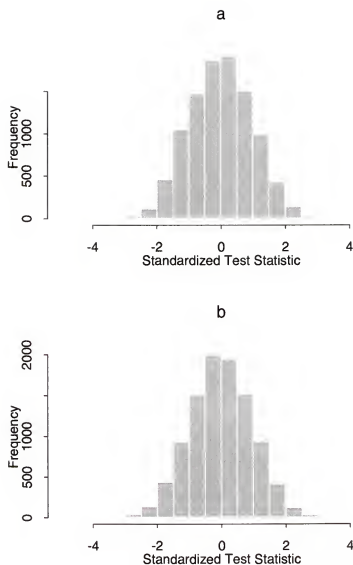


Figure 5-10. Standardized test statistics for parameter setting 2 for 10000 random samples with size 1000. a) Mixture model; b) Three-piece Gompertz model.

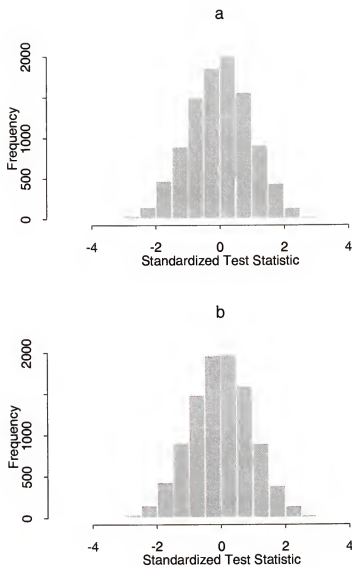


Figure 5-11. Standardized test statistics for parameter setting 2 for 10000 random samples with size 2000. a) Mixture model; b) Three-piece Gompertz model.

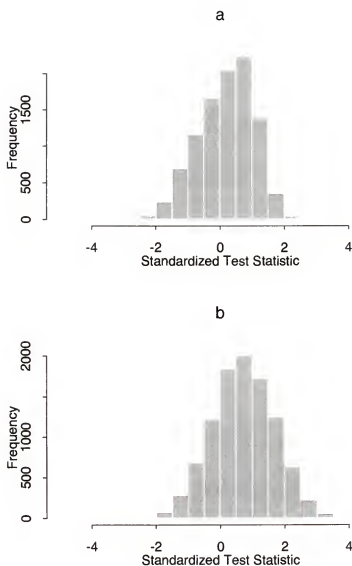


Figure 5-12. Standardized test statistics for parameter setting 3 for 10000 random samples with size 400. a) Mixture model; b) Three-Piece Gompertz model.

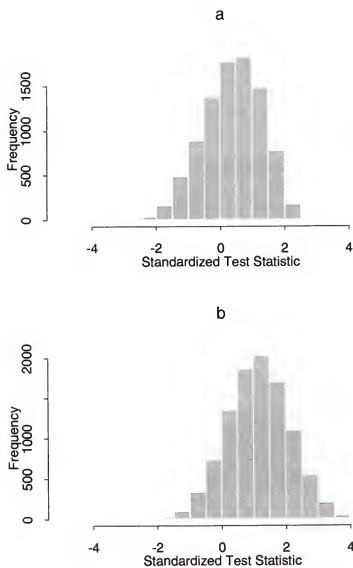


Figure 5-13. Standardized test statistics for parameter setting 3 for 10000 random samples with size 1000. a) Mixture model; b) Three-piece Gompertz model.

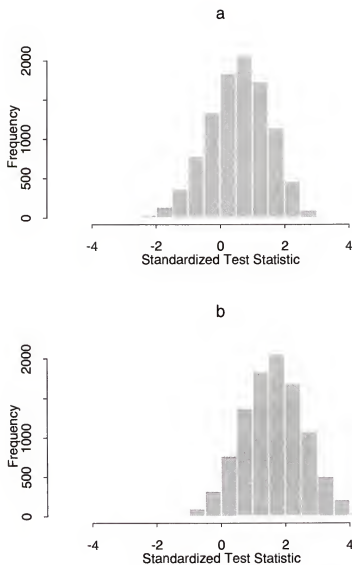


Figure 5-14. Standardized test statistics for parameter setting 3 for 10000 random samples with size 2000. a) Mixture model; b) Three-piece Gompertz model.

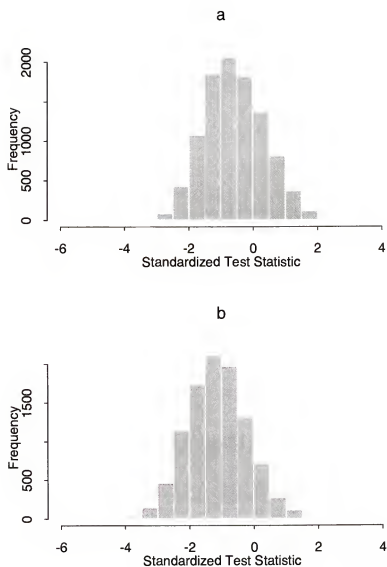


Figure 5-15. Standardized test statistics for parameter setting 4 for 10000 random samples with size 400. a) Mixture model; b) Three-piece Gompertz model.

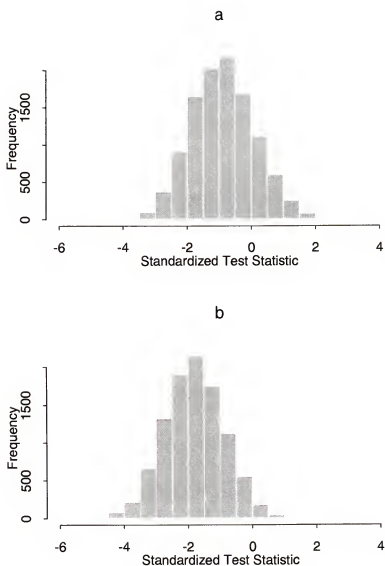


Figure 5-16. Standardized test statistics for parameter setting 4 for 10000 random samples with size 1000. a) Mixture model; b) Three-piece Gompertz model.

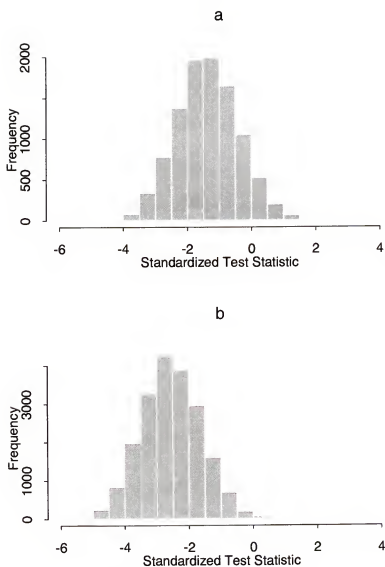


Figure 5-17. Standardized test statistics for parameter setting 4 for 10000 random samples with size 2000. a) Mixture model; b) Three-piece Gompertz model.

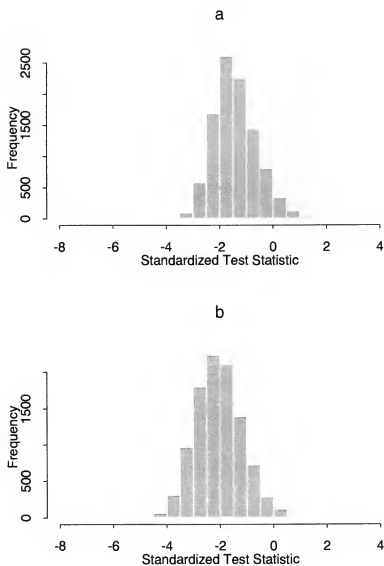


Figure 5-18. Standardized test statistics for parameter setting 5 for 10000 random samples with size 400. a) Mixture model; b) Three-piece Gompertz model.

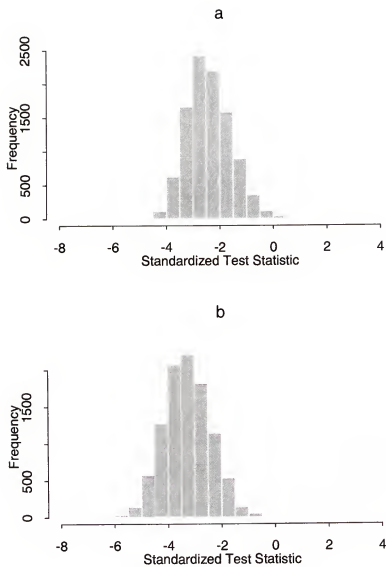


Figure 5-19. Standardized test statistics for parameter setting 5 for 10000 random samples with size 1000. a) Mixture model; b) Three-piece Gompertz model.

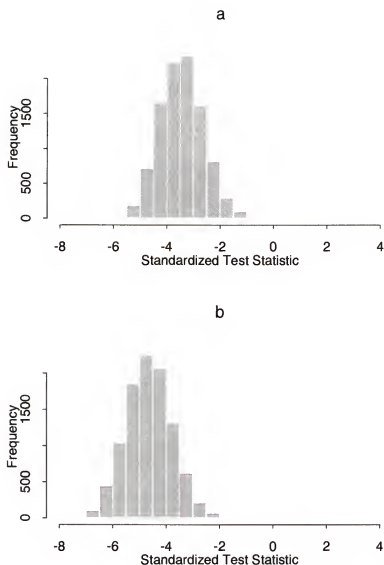


Figure 5-20. Standardized test statistics for parameter setting 5 for 10000 random samples with size 2000. a) Mixture model; b) Three-piece Gompertz model.

CHAPTER 6

EXAMPLE

6.1 Introduction

In Chapter 2 we proposed a test statistic that addresses the hypothesis presented in Equation (2.3). Subsequently in Chapter 3 and Chapter 4 we established the asymptotic normality of parameter estimates in the piecewise Gompertz model. The asymptotic normality validated the test of hypothesis proposed in Chapter 2. Chapter 5 reported Monte Carlo Type I error and power calculations, which established the strength of our test statistic on the comparison of cure rates under various parameter settings. In this chapter, we demonstrate the applicability of our test statistic with a data set.

The data set comes from a study by the Pediatric Oncology Group (POG) at the university of Florida (Ravindranath et al. 1996). The motivation of this study is to compare cure proportions among patients with acute myeloid leukemia (AML) under different treatments. Three treatments, allogeneic bone marrow transplantation (allo. BMT), autologous bone marrow transplantation (auto. BMT), and intensive chemotherapy, are performed in the study. In order to register in the study, all patients are required to be in remission after the first two courses of induced chemotherapy. The POG evaluated 666 patients whose ages were under 21 from June 1988 to March 1993. A total of 552 patients entered remission and are therefore eligible for the study. Of the eligible patients, 89 patients received allo. BMT; 117 were assigned to an additional six-course intensive chemotherapy; and 115 were assigned to receive auto. BMT. There were 231 patients who did not participate in the study due to various reasons.

6.2 Methodology

Our objective based on this data set is to perform pairwise comparisons on cure rates between treatment groups after controlling for patient's gender. We compare tests for three parametric models and for a proportion test based on KM estimates of cure. The three parametric models are the logistic-exponential mixture model, the one-piece Gompertz model, and the three-piece Gompertz model. In addition, we present cure rates and their standard deviations for every treatment and gender combination based on estimates from parametric models and on KM estimates. In the end, we compare the standard one-piece Gompertz model with the three-piece Gompertz model by applying the goodness-of-fit test we derived in Chapter 5. In this section, we will briefly describe 1) three parametric models and their corresponding test statistics; and 2) the proportion test based on KM estimates of cure proportions.

In the exponential-logistic mixture model, we assume an exponential distribution with parameter λ for the failure time and a logistic regression model for the cure proportion. The expression $\log(\lambda)$ is defined as a linear function of the covariate x and the treatment indicator. We define two dummy variables to represent the three-level treatment indicator. They are presented as $G_1 = I_{[\text{if patient in treatment 1}]}$ and $G_2 = I_{[\text{if patient in treatment 2}]}$. The survival function for the failure time given that the observation is not a long-term survivor ($Z = 1$) is

$$S^*(t|Z = 1) = e^{-\lambda t} = \exp\{-te^{\gamma_0 + \gamma_1 x_1 + \gamma_2 g_1 + \gamma_3 g_2}\}. \quad (6.1)$$

The cure probability is defined as

$$\pi_g(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 g_1 + \beta_3 g_2}}. \quad (6.2)$$

Therefore the hypothesis of testing the equality of cure rates in three treatment groups can be written as

$$H_0 : \beta_2 = \beta_3 = 0. \quad (6.3)$$

Pairwise comparisons between two treatments are based on the following standardized test statistics:

$$T_1 = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{V(\hat{\beta}_2 - \hat{\beta}_3)}}, \quad (6.4)$$

$$T_2 = \frac{\hat{\beta}_2}{\sqrt{V(\hat{\beta}_2)}}, \text{ and} \quad (6.5)$$

$$T_3 = \frac{\hat{\beta}_3}{\sqrt{V(\hat{\beta}_3)}}. \quad (6.6)$$

Note that T_1 is the test statistic that compares cure proportions between Treatments 1 and 2; T_2 compares cure proportions between Treatments 1 and 3; and T_3 compares cure proportions between Treatments 2 and 3. The variances are obtained by inverting the observed information matrix obtained by taking the second derivatives of the log-likelihood function with respect to parameters.

The one-piece Gompertz model is a special case of the piecewise Gompertz model. Let λ_g be the parameter associated with the treatment g and β_1 the coefficient associated with the covariate x . For $\lambda_g < 0$, the cure proportion for patients with the treatment g and the covariate x is written as

$$\pi_g(x) = \exp\left\{e^{\beta_0 + \beta_1 x} \frac{1}{\lambda_g}\right\}. \quad (6.7)$$

If $\hat{\lambda}_g < 0$, there is cure proportion for patients under the g^{th} treatment. If $\lambda_g < 0$ for any g , then the hypothesis of testing the equality of cure proportions is written as

$$H_0 : \lambda_1 = \lambda_2 = \lambda_3. \quad (6.8)$$

Standardized test statistics for pairwise comparisons of cure proportions are presented as

$$T_1 = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\sqrt{V(\hat{\lambda}_1 - \hat{\lambda}_2)}}, \quad (6.9)$$

$$T_2 = \frac{\hat{\lambda}_1 - \hat{\lambda}_3}{\sqrt{V(\hat{\lambda}_1 - \hat{\lambda}_3)}}, \text{ and} \quad (6.10)$$

$$T_3 = \frac{\hat{\lambda}_2 - \hat{\lambda}_3}{\sqrt{V(\hat{\lambda}_2 - \hat{\lambda}_3)}}. \quad (6.11)$$

Similarly these three standardized test statistics are for the comparison of cure proportions between Treatments 1 and 2, between Treatments 1 and 3, and between Treatments 2 and 3, respectively. The variances are obtained from the inverse of the observed information matrix.

Finally, as we have discussed in Chapter 2, the cure proportion in a piecewise Gompertz model as a function of the covariate x and the treatment g is expressed in Equation (2.4). We consider a three-interval case in this example. Denote the three intervals by $[\tau_0, \tau_1)$, $[\tau_1, \tau_2)$, and $[\tau_2, \tau_3)$, where $\tau_0 = 0$ and $\tau_3 = \infty$. When $\lambda_{3g} < 0$ for any g , testing the equality of cure proportions for any given value of x is equivalent to testing

$$H_0 : \sum_{i=1}^3 \frac{e^{\lambda_{i1}\tau_{i-1}} - e^{\lambda_{i1}\tau_i}}{\lambda_{i1}} = \sum_{i=1}^3 \frac{e^{\lambda_{i2}\tau_{i-1}} - e^{\lambda_{i2}\tau_i}}{\lambda_{i2}} = \sum_{i=1}^3 \frac{e^{\lambda_{i3}\tau_{i-1}} - e^{\lambda_{i3}\tau_i}}{\lambda_{i3}}. \quad (6.12)$$

Standardized test statistics for pairwise comparisons between treatment groups are presented as follows:

$$T_1 = \frac{\sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i1}\tau_{i-1}} - e^{\hat{\lambda}_{i1}\tau_i}}{\hat{\lambda}_{i1}} - \sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i2}\tau_{i-1}} - e^{\hat{\lambda}_{i2}\tau_i}}{\hat{\lambda}_{i2}}}{\sqrt{V(\sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i1}\tau_{i-1}} - e^{\hat{\lambda}_{i1}\tau_i}}{\hat{\lambda}_{i1}} - \sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i2}\tau_{i-1}} - e^{\hat{\lambda}_{i2}\tau_i}}{\hat{\lambda}_{i2}})}}, \quad (6.13)$$

$$T_2 = \frac{\sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i1}\tau_{i-1}} - e^{\hat{\lambda}_{i1}\tau_i}}{\hat{\lambda}_{i1}} - \sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i3}\tau_{i-1}} - e^{\hat{\lambda}_{i3}\tau_i}}{\hat{\lambda}_{i3}}}{\sqrt{V(\sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i1}\tau_{i-1}} - e^{\hat{\lambda}_{i1}\tau_i}}{\hat{\lambda}_{i1}} - \sum_{i=1}^3 \frac{e^{\hat{\lambda}_{i3}\tau_{i-1}} - e^{\hat{\lambda}_{i3}\tau_i}}{\hat{\lambda}_{i3}})}}, \text{ and} \quad (6.14)$$

$$T_3 = \frac{\sum_{i=1}^3 \frac{e^{\widehat{\lambda}_{i2}\tau_i-1} - e^{\widehat{\lambda}_{i2}\tau_i}}{\widehat{\lambda}_{i2}} - \sum_{i=1}^3 \frac{e^{\widehat{\lambda}_{i3}\tau_i-1} - e^{\widehat{\lambda}_{i3}\tau_i}}{\widehat{\lambda}_{i3}}}{\sqrt{V(\sum_{i=1}^3 \frac{e^{\widehat{\lambda}_{i2}\tau_i-1} - e^{\widehat{\lambda}_{i2}\tau_i}}{\widehat{\lambda}_{i2}} - \sum_{i=1}^3 \frac{e^{\widehat{\lambda}_{i3}\tau_i-1} - e^{\widehat{\lambda}_{i3}\tau_i}}{\widehat{\lambda}_{i3}})}}. \quad (6.15)$$

These test statistics T_1 , T_2 , and T_3 are for comparisons of cure proportions between Treatments 1 and 2, between Treatments 1 and 3, and between Treatments 2 and 3, respectively. Variance estimates are not quite straightforward as for tests based on the mixture model and on the one-piece Gompertz model. They have already been presented in Chapter 2 and will not be repeated here.

Proportion tests of KM estimates of cure for pairwise comparisons between treatments are generated based on KM estimates of cure proportions and standard errors of the estimated cure proportions.

P-values for tests derived from parametric models and for the proportion test will be presented in the next section. In addition, we present survival curves estimated based on both parametric models and KM estimates. Furthermore, we present the cure proportion and the corresponding standard deviation for each treatment and gender combination based on parametric estimates and on KM estimates. The result of comparing the one-piece Gompertz model with the three-piece model by the goodness-of-fit test we derived earlier is also presented.

6.3 Results

6.3.1 Pairwise Comparisons on Cure Proportions

In Table (6-1), we present p-values for tests based on three parametric models and for the proportion test based on KM estimates. The three intervals for the three-piece Gompertz model are defined as $[0, 50)$, $[50, 140)$, and $[140, \infty)$. All numbers presented here are in terms of weeks in follow-up. Tests based on the exponential-logistic mixture model, the three-piece Gompertz model, and the KM estimates detect the difference in cure proportions between allo. BMT and auto.

BMT and the difference between allo. BMT and intensive chemotherapy. The allo. BMT increases cure proportion significantly more than the other two treatments. There is no significant difference between auto. BMT and chemotherapy. A test based on the one-piece Gompertz model does not show any significant difference between any two of these three treatments. The maximization procedures obtaining MLEs of parameters for both the one-piece Gompertz model and the three-piece Gompertz model are smooth. The algorithm converges within a few steps. However, there is a convergence problem related to the mixture model. Different starting values for parameters are chosen in order to obtain the convergence. We have discussed the convergence problem related to the mixture model in both Chapter 1 and Chapter 5. The convergence problem we encounter here confirms our earlier assertion.

Table 6-1. Comparison of Tests from the Mixture Model, the One-Piece Gompertz Model, and the Three-Piece Gompertz Model, and the Proportion Test on Cure Proportions.

Comparisons	P-values for Treatment Comparisons			
	Mixture	1-Piece Gompertz	3-Piece Gompertz	Proportion Test
Allo. BMT vs Auto. BMT	0.016	0.384	0.015	0.011
Allo. BMT vs chemotherapy	0.012	0.101	0.003	0.007
Auto. BMT vs chemotherapy	0.690	0.516	0.897	0.741

In Figures (6-1) to (6-4), we show estimated survival curves from different parametric models and from KM estimates. The three curves on Figure (6-1)a are survival curves for three treatment groups for male patients based on the mixture model. Figure (6-1)b presents three treatment survival curves for female patients based on the mixture model. Figure (6-2) presents survival curves estimated from the one-piece Gompertz model for male and female patients in three treatment groups. Survival curves estimated from the three-piece Gompertz model are

presented in Figure (6-3). The KM estimates of survival curves are presented in Figure (6-4).

6.3.2 Cure Proportion Estimates

Estimates of cure proportions are presented in Table (6-2). We present estimates of cure rates and their standard deviations from the mixture model, the one-piece Gompertz model, the three-piece Gompertz model, and the KM estimate for all combinations of patient's gender and three treatment groups. Treatment 1 represents allo. BMT, treatment 2 represents auto. BMT, and Treatment 3 represents intensive chemotherapy. Note that cure proportion exists for each of the combinations. Female patients show higher cure probability than male patients. Standard errors for cure proportions estimated from both the three-piece Gompertz model and the mixture model are smaller compared to the one-piece Gompertz model and the KM estimates.

Table 6-2. Estimates and Standard Deviations of Cure Proportion Generated from the Mixture Model, the One-Piece Gompertz Model, the Three-Piece Gompertz Model, and the Kaplan-Meier Estimates.

Gender	Treatment	Cure Proportions (Standard Deviations)			
		Mixture	1-piece Gompertz	3-piece Gompertz	KM
M	1	0.558 (0.059)	0.504 (0.062)	0.568 (0.056)	0.648 (0.076)
M	2	0.386 (0.050)	0.445 (0.060)	0.397 (0.050)	0.387 (0.065)
M	3	0.407 (0.033)	0.409 (0.050)	0.405 (0.033)	0.388 (0.036)
F	1	0.597 (0.056)	0.533 (0.062)	0.596 (0.054)	0.531 (0.071)
F	2	0.424 (0.051)	0.475 (0.061)	0.429 (0.051)	0.421 (0.065)
F	3	0.445 (0.036)	0.440 (0.051)	0.437 (0.035)	0.466 (0.041)

6.3.3 Goodness-of-Fit

From the above discussion, we observe that the three-piece Gompertz model fits better than the one-piece Gompertz model. It is natural to compare these two models by applying the goodness-of-fit test we derived in this dissertation. The null

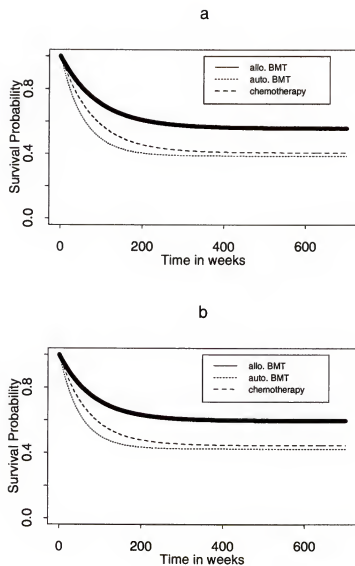


Figure 6-1. Survival curves estimated from the mixture model for three treatments.
a) Male; b) Female.

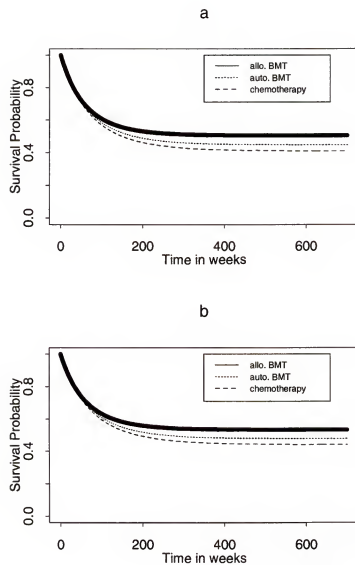


Figure 6-2. Survival curves estimated from the one-piece Gompertz model for three treatments. a) Male; b) Female.

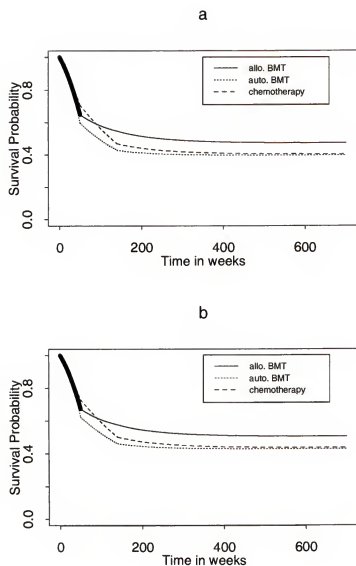


Figure 6-3. Survival curves estimated from the three-piece Gompertz model for three treatments. a) Male; b) Female.

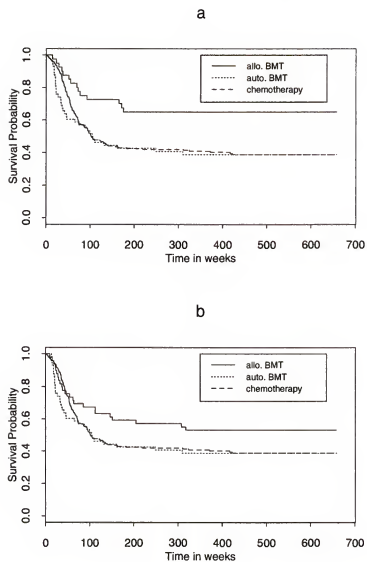


Figure 6-4. Survival curves from Kaplan-Meier estimates for three treatments. a) Male; b) Female.

hypothesis is expressed as

$$H_0 : \lambda_{1g} = \lambda_{2g} = \lambda_{3g},$$

for any $g = 1, 2$, and 3 . Note that λ_{2g} and λ_{3g} are parameters for the second and third interval in the three-piece Gompertz model, respectively. We showed that under H_0

$$-2(l_1 - l_2) \stackrel{D}{\sim} \chi_6^2,$$

where l_1 and l_2 are the log-likelihood functions for the one-piece and the three-piece Gompertz models, respectively. From the computation, we obtain that $l_1 = -1977.289$ and $l_2 = -1950.979$. Therefore, the test is highly significant. It implies that the three-piece model is significantly better than the one-piece Gompertz model.

CHAPTER 7

SUMMARY

In this dissertation, we presented a piecewise Gompertz model to address long-term survival problem. We constructed a standardized test statistic based on our model which compares cure rates between treatment groups. We derived the strict concavity of the likelihood function based on the model. In addition, we derived the asymptotic properties of the MLEs of parameters in the model. A goodness-of-fit test was also derived to access an appropriate number of intervals for the model.

In oncology study when there are long-term survivors and when there are different hazard patterns on different time intervals, one might consider using our piecewise Gompertz model. Our model is more computational feasible than both the mixture model and the semi-parametric model with bounded cumulative hazard function. In addition, our model is more flexible due to its piecewise structure. Our model also provides a reliable estimate of the cure rate than the KM estimate when the sample size is small at the end of the study period.

In the future, we plan to find an optimal partition of the time span by deriving a test to compare different partitions. Sample size determination will also be assessed. Since we notice that there are jumps on the survival curve defined by the piecewise Gompertz model, we might consider using a log-spline model to achieve smoothness of the curve. In addition, we plan to perform more detailed simulation studies on the semiparametric model which assumes bounded cumulative hazard function and to study the asymptotic behavior of the semiparametric model.

APPENDIX A
DERIVATION OF MARGINAL LIKELIHOOD

Let

$$I_n = \int_{t_0=0}^{\infty} \cdots \int_{t_{n-1}}^{\infty} \prod_{i=1}^n f(t_i) e^{-F(t_i)\Theta_i} dt_n dt_{n-1} \cdots dt_1$$

and define

$$I_k = \int_{t_{n-k}}^{\infty} \cdots \int_{t_{n-1}}^{\infty} \prod_{i=n-k+1}^n f(t_i) e^{-F(t_i)\Theta_i} dt_n dt_{n-1} \cdots dt_{n-k+1}, \quad k = 1, 2, \dots, n.$$

As $k = 1$,

$$I_1 = \int_{t_{n-1}}^{\infty} f(t_n) e^{-F(t_n)\Theta_n} dt_n = \frac{1}{\Theta_n} [e^{-F(t_{n-1})\Theta_n} - e^{-\Theta_n}], \quad (\text{A.1})$$

and when $k = 2$,

$$\begin{aligned} I_2 &= \int_{t_{n-2}}^{\infty} \int_{t_{n-1}}^{\infty} \prod_{i=n-1}^n f(t_i) e^{-F(t_i)\Theta_i} dt_n dt_{n-1} \\ &= \int_{t_{n-2}}^{\infty} f(t_{n-1}) e^{-F(t_{n-1})\Theta_{n-1}} \left\{ \int_{t_{n-1}}^{\infty} f(t_n) e^{-F(t_n)\Theta_n} dt_n \right\} dt_{n-1} \\ &= \int_{t_{n-2}}^{\infty} f(t_{n-1}) e^{-F(t_{n-1})\Theta_{n-1}} \frac{1}{\Theta_n} \{e^{-F(t_{n-1})\Theta_n} - e^{-\Theta_n}\} dt_{n-1} \\ &= \frac{1}{\Theta_n(\Theta_{n-1} + \Theta_n)} [e^{-(\Theta_{n-1} + \Theta_n)F(t_{n-2})} - e^{-(\Theta_{n-1} + \Theta_n)}] \\ &\quad - \frac{e^{-\Theta_n}}{\Theta_n \Theta_{n-1}} [e^{-\Theta_{n-1}F(t_{n-2})} - e^{-\Theta_{n-1}}]. \end{aligned}$$

By similar derivation, we obtain that when $k = 3$,

$$\begin{aligned} I_3 &= \int_{t_{n-3}}^{\infty} f(t_{n-2}) e^{-F(t_{n-2})\Theta_{n-2}} I_2 dt_{n-2} \\ &= \frac{1}{\Theta_n(\Theta_n + \Theta_{n-1})(\Theta_n + \Theta_{n-1} + \Theta_{n-2})} [e^{-(\Theta_n + \Theta_{n-1} + \Theta_{n-2})F(t_{n-3})} - e^{-(\Theta_n + \Theta_{n-1} + \Theta_{n-2})}] \end{aligned}$$

$$\begin{aligned}
& -\frac{e^{-\Theta_n}}{\Theta_n \Theta_{n-1} (\Theta_{n-1} + \Theta_{n-2})} [e^{-(\Theta_{n-1} + \Theta_{n-2})F(t_{n-3})} - e^{-(\Theta_{n-1} + \Theta_{n-2})}] \\
& + \frac{e^{-(\Theta_n + \Theta_{n-1})}}{\Theta_{n-1} \Theta_{n-2} (\Theta_n + \Theta_{n-1})} [e^{-\Theta_{n-2}F(t_{n-3})} - e^{-\Theta_{n-2}}].
\end{aligned}$$

Let

$$I_k = \sum_{i=1}^k (-1)^{i-1} A_{i,k} [e^{-S_{n-k+1, n-i+1}F(t_{n-k})} - e^{-S_{n-k+1, n-i+1}}], \quad (\text{A.2})$$

then

$$\begin{aligned}
I_{k+1} &= \int_{t_{n-k-1}}^{\infty} I_k f(t_{n-k}) e^{-F(t_{n-k})\Theta_{n-k}} dt_{n-k} \\
&= \int_{t_{n-k-1}}^{\infty} \sum_{i=1}^k (-1)^{i-1} A_{i,k} [e^{-S_{n-k+1, n-i+1}F(t_{n-k})} - e^{-S_{n-k+1, n-i+1}}] \\
&\quad f(t_{n-k}) e^{-F(t_{n-k})\Theta_{n-k}} dt_{n-k} \\
&= \sum_{i=1}^k (-1)^{i-1} A_{i,k} \frac{1}{S_{n-k, n-i+1}} [e^{-S_{n-k, n-i+1}F(t_{n-k-1})} - e^{-S_{n-k, n-i+1}}] \\
&\quad + \sum_{i=1}^k (-1)^i \frac{1}{\Theta_{n-k}} A_{i,k} e^{-S_{n-k+1, n-i+1}} [e^{-F(t_{n-k-1})\Theta_{n-k}} - e^{-\Theta_{n-k}}]. \quad (\text{A.3})
\end{aligned}$$

Also we know that

$$\begin{aligned}
I_{k+1} &= \sum_{i=1}^{k+1} (-1)^{i-1} A_{i,k+1} [e^{-S_{n-k, n-i+1}F(t_{n-k-1})} - e^{-S_{n-k, n-i+1}}] \\
&= \sum_{i=1}^k (-1)^{i-1} A_{i,k+1} [e^{-S_{n-k, n-i+1}F(t_{n-k-1})} - e^{-S_{n-k, n-i+1}}] \\
&\quad + (-1)^k A_{k+1, k+1} [e^{-S_{n-k, n-k}F(t_{n-k-1})} - e^{-S_{n-k, n-k}}]. \quad (\text{A.4})
\end{aligned}$$

Equate (A.3) and (A.4), we obtain that

$$A_{i,k+1} = A_{i,k} \frac{1}{S_{n-k, n-i+1}} \quad (\text{A.5})$$

and

$$A_{k+1, k+1} = \sum_{i=1}^k (-1)^{i-k} \frac{A_{i,k}}{\Theta_{n-k}} e^{-S_{n-k+1, n-i+1}}. \quad (\text{A.6})$$

Therefore by Equation (A.5)

$$\begin{aligned} A_{i,k} &= A_{i,k-1} \frac{1}{S_{n-k+1,n-i+1} S_{n-k,n-i+1}} \\ &= A_{i,i} \frac{1}{S_{n-i,n-i+1} S_{n-i-1,n-i+1} \cdots S_{n-k+1,n-i+1}} = \frac{A_{i,i}}{P_{n-k+1,n-i+1}} \end{aligned} \quad (\text{A.7})$$

for $i \leq k$ and by the definition of $P_{a,b}$ given in Chapter 1. In addition by Equation (A.6),

$$\begin{aligned} A_{k+1,k+1} &= \sum_{i=1}^k (-1)^{i-k} \frac{A_{i,k}}{\Theta_{n-k}} e^{-S_{n-k+1,n-i+1}} \\ &= \sum_{i=1}^{k-1} (-1)^{i-k} \frac{A_{i,k}}{\Theta_{n-k}} e^{-S_{n-k+1,n-i+1}} + \frac{A_{k,k}}{\Theta_{n-k}} e^{-S_{n-k+1,n-k+1}} \\ &= \sum_{i=1}^{k-1} (-1)^{i-k} A_{i,i} \frac{1}{S_{n-i,n-i+1} S_{n-i-1,n-i+1} \cdots S_{n-k+1,n-i+1}} \frac{e^{-S_{n-k+1,n-i+1}}}{S_{n-k,n-k}} \\ &\quad + A_{k,k} \frac{e^{-S_{n-k+1,n-k+1}}}{\Theta_{n-k}}. \end{aligned} \quad (\text{A.8})$$

Also we observe that

$$I_1 = A_{1,1} [e^{-S_{n,n} F(t_{n-1})} - e^{-S_{n,n}}] = \frac{1}{\Theta_n} [e^{-S_{n,n} F(t_{n-1})} - e^{-S_{n,n}}],$$

where the first equality is true by Equation (A.2) and the second is true from Equation (A.1). Therefore

$$A_{1,1} = \frac{1}{S_{n,n}} = \frac{1}{\Theta_n}.$$

Similarly we obtain through derived Equation (A.8)

$$A_{2,2} = A_{11} \frac{\Theta_{n-1}}{e^{-S_{n,n}}} = \frac{e^{-S_{n,n}}}{S_{n,n} S_{n-1,n-1}},$$

and

$$A_{3,3} = \frac{e^{-S_{n-1,n}}}{S_{n-1,n-1} S_{n-2,n-2} S_{n-1,n}}.$$

If we let

$$A_{k,k} = \frac{e^{-S_{n-k+2,n}}}{S_{n-k+1,n-k+1} P^{n-k+2,n}}, \quad (\text{A.9})$$

where $P^{n-k+2,n}$ was defined earlier in Chapter 1 as $\prod_{j=n-k+2}^n S_{n-k+2,j}$. In addition, we observe from Equation (A.8) that

$$\begin{aligned}
 & A_{k,k} \\
 &= \frac{1}{S_{n-k+1,n-k+1}} \sum_{i=1}^{k-2} (-1)^{i-k+1} A_{i,i} \frac{S_{n-i+1,n-i+1} e^{-S_{n-k+2,n-i+1}}}{P^{n-k+2,n-i+1}} \\
 &\quad + A_{k-1,k-1} e^{-S_{n-k+2,n-i+1}} \\
 &= \frac{e^{-S_{n-k+2,n}}}{S_{n-k+1,n-k+1}} \left\{ \sum_{i=1}^{k-1} (-1)^{i-k+1} \frac{1}{P^{n-i+2,n} P^{n-k+2,n-i+1}} \right\}. \tag{A.10}
 \end{aligned}$$

Hence from equating Equations (A.9) and (A.10), the following equation holds.

$$\sum_{i=0}^{k-1} (-1)^i \frac{1}{P^{n-i+1,n} P^{n-k+1,n-i}} = (-1)^{k-1} \frac{1}{P^{n-k+1,n}}. \tag{A.11}$$

Now we can derive $A_{k+1,k+1}$ as follows:

$$\begin{aligned}
 & A_{k+1,k+1} \\
 &= \frac{1}{S_{n-k,n-k}} \left\{ \sum_{i=1}^{k-1} (-1)^{i-k} A_{i,i} \frac{S_{n-i+1,n-i+1} e^{-S_{n-k+1,n-i+1}}}{P^{n-k+1,n-i+1}} + A_{k,k} e^{-S_{n-k+1,n-k+1}} \right\} \\
 &= \frac{e^{-S_{n-k+2,n}}}{S_{n-k,n-k}} \left\{ \sum_{i=1}^{k-1} (-1)^{i-k} \frac{1}{P^{n-i+2,n} P^{n-k+1,n-i+1}} + \frac{1}{S_{n-k+1,n-k+1} P^{n-k+2,n}} \right\} \\
 &= \frac{e^{-S_{n-k+2,n}}}{S_{n-k,n-k}} \left\{ \frac{1}{P^{n-k+1,n}} - \frac{1}{P^{n-k+2,n} P^{n-k+1,n-k+1}} + \frac{1}{S_{n-k+1,n-k+1} P^{n-k+2,n}} \right\} \\
 &= \frac{e^{-S_{n-k+2,n}}}{S_{n-k,n-k} P^{n-k+1,n}}, \tag{A.12}
 \end{aligned}$$

by Equation (A.11). Therefore,

$$A_{i,k} = A_{i,i} \frac{S_{n-i+1,n-i+1}}{P_{n-k+1,n-i+1} P^{n-i+2,n}}, \tag{A.13}$$

for $i \leq k$. I_k , as defined in Equation (A.2), therefore can be written as

$$I_k = \sum_{i=1}^k (-1)^{i-1} \frac{e^{-S_{n-i+2,n} - S_{n-k+1,n-i+1} F(t_{n-k})} - e^{-S_{n-k+1,n}}}{P^{n-i+2,n} P_{n-k+1,n-i+1}}. \tag{A.14}$$

Set $k = n$, we obtain

$$\begin{aligned}
& I_n \\
&= \sum_{i=1}^n (-1)^{i-1} \frac{1}{P^{n-i+2,n} P_{1,n-i+1}} [e^{-S_{n-i+2,n} - S_{1,n-i+1} F(t_0)} - e^{-S_{1,n}}] \\
&= \sum_{i=1}^n (-1)^{i-1} \frac{1}{P^{n-i+2,n} P_{1,n-i+1}} [e^{-S_{n-i+2,n}} - e^{-S_{1,n}}] \\
&= \sum_{i=1}^n (-1)^{i-1} \frac{e^{-S_{n-i+2,n}}}{P^{n-i+2,n} P_{1,n-i+1}} - \sum_{i=1}^n (-1)^{i-1} \frac{e^{-S_{1,n}}}{P^{n-i+2,n} P_{1,n-i+1}} \\
&= \sum_{i=0}^{n-1} (-1)^i \frac{e^{-S_{n-i+1,n}}}{P^{n-i+1,n} P_{1,n-i}} - \sum_{i=0}^{n-1} (-1)^i \frac{e^{-S_{1,n}}}{P^{n-i+1,n} P_{1,n-i}} \\
&= \sum_{i=0}^{n-1} (-1)^i \frac{e^{-S_{n-i+1,n}}}{P^{n-i+1,n} P_{1,n-i}} + (-1)^n \frac{e^{-S_{1,n}}}{P^{1,n}} \\
&= \sum_{i=0}^n (-1)^i \frac{e^{-S_{n-i+1,n}}}{P^{n-i+1,n} P_{1,n-i}}, \tag{A.15}
\end{aligned}$$

since

$$\sum_{i=0}^{n-1} \frac{e^{-S_{1,n}}}{P_{1,n-i} P^{n-i+1,n}} = (-1)^{n-1} \frac{e^{-S_{1,n}}}{P^{1,n}},$$

by Equation (A.11). Hence the marginal likelihood function defined in Equation (1.14) can be derived since we observe that

$$L_m = \int_{t_0=0}^{\infty} \cdots \int_{t_{n-1}}^{\infty} L = \prod_{i=1}^n \theta_i^d I_n,$$

where L was defined in Equation (1.13).

APPENDIX B PROOF OF LEMMA 4.1

The results presented in Lemma 4.1 hold for any general parametric model if it satisfies certain regularity conditions. In our proof, we show that the results also hold for any parametric model with independent censoring under certain regularity conditions. Let $T^* = \min(T, U)$ be the observed survival/censoring time and Δ be the censoring indicator. Denote the hazard function, the probability density function, and the survival function for the survival time T by $h(T; \boldsymbol{\theta})$, $f(T; \boldsymbol{\theta})$, and $S(T; \boldsymbol{\theta})$, respectively. Denote the survival function for the censoring time U by $G_U(u)$. Then the likelihood function is written as

$$L(T^*, \Delta) = h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta}).$$

Therefore the expectation of the score function $S(\boldsymbol{\theta})$ has the following expression:

$$\begin{aligned} & E[S(\boldsymbol{\theta})] \\ &= E\left[\frac{d}{d\boldsymbol{\theta}} \log L(T^*, \Delta)\right] = E\left[\frac{\frac{d}{d\boldsymbol{\theta}} L(T^*, \Delta)}{L(T^*, \Delta)}\right] \\ &= E\left[\frac{\frac{d}{d\boldsymbol{\theta}} (h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta}))}{h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta})} I(\Delta = 1)\right] + E\left[\frac{\frac{d}{d\boldsymbol{\theta}} (h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta}))}{h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta})} I(\Delta = 0)\right]. \end{aligned} \tag{B.1}$$

We can write the first term on the right-hand side of Equation (B.1) as

$$\begin{aligned} & E\left[\frac{\frac{d}{d\boldsymbol{\theta}} (h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta}))}{h^\Delta(T^*; \boldsymbol{\theta}) S(T^*; \boldsymbol{\theta})} I(\Delta = 1)\right] = E\left[\frac{\frac{d}{d\boldsymbol{\theta}} f(T; \boldsymbol{\theta})}{f(T; \boldsymbol{\theta})} I(T \leq U)\right] \\ &= \int_0^\infty \frac{\frac{d}{d\boldsymbol{\theta}} f(u; \boldsymbol{\theta})}{f(u; \boldsymbol{\theta})} f(u; \boldsymbol{\theta}) G_U(u) du = \frac{d}{d\boldsymbol{\theta}} \int_0^\infty f(u; \boldsymbol{\theta}) G_U(u) du \\ &= \frac{d}{d\boldsymbol{\theta}} P(T \leq U). \end{aligned}$$

Similarly, the second term of the right-hand side of Equation (B.1) can be written as

$$\begin{aligned}
& E\left[\frac{\frac{d}{d\boldsymbol{\theta}}(h^\Delta(T^*; \boldsymbol{\theta})S(T^*; \boldsymbol{\theta}))}{h^\Delta(T^*; \boldsymbol{\theta})S(T^*; \boldsymbol{\theta})}I(\Delta = 0)\right] = E\left[\int_0^\infty \frac{\frac{d}{d\boldsymbol{\theta}}S(U; \boldsymbol{\theta})}{S(U; \boldsymbol{\theta})}I(T \geq U)\right] \\
& = \int_0^\infty \frac{\frac{d}{d\boldsymbol{\theta}}S(u; \boldsymbol{\theta})}{S(u; \boldsymbol{\theta})}S(u; \boldsymbol{\theta})d(1 - G_U(u)) = \frac{d}{d\boldsymbol{\theta}} \int_0^\infty S(u; \boldsymbol{\theta})d(1 - G_U(u)) \\
& = \frac{d}{d\boldsymbol{\theta}}P(T > U).
\end{aligned}$$

Hence,

$$E[S(\boldsymbol{\theta})] = \frac{d}{d\boldsymbol{\theta}}[P(T \leq U) + P(T > U)] = \mathbf{0}.$$

Now we need to prove $V[S(\boldsymbol{\theta})] = D$. Let $\frac{d^2 L}{d\boldsymbol{\theta}^2}$, $\frac{d^2 \log L}{d\boldsymbol{\theta}^2}$, and $\frac{d}{d\boldsymbol{\theta}}(\frac{d}{d\boldsymbol{\theta}}L)$ represent the $(p + IG) \times (p + IG)$ dimensional matrices formed by corresponding second order derivatives. Note that

$$\begin{aligned}
\frac{d^2 \log L}{d\boldsymbol{\theta}^2} &= \frac{d}{d\boldsymbol{\theta}}\left(\frac{\frac{d}{d\boldsymbol{\theta}}L}{L}\right) = \frac{L \frac{d^2}{d\boldsymbol{\theta}^2}L - (\frac{d}{d\boldsymbol{\theta}}L)(\frac{d}{d\boldsymbol{\theta}}L)^T}{L^2} \\
&= \frac{1}{L}\left(\frac{d^2 L}{d\boldsymbol{\theta}^2}\right) - \left(\frac{d}{d\boldsymbol{\theta}}\log L\right)\left(\frac{d}{d\boldsymbol{\theta}}\log L\right)^T.
\end{aligned}$$

It is clear that $V[S(\boldsymbol{\theta})] = D(\boldsymbol{\theta})$ if we can prove

$$E\left[\frac{1}{L} \frac{d^2 L}{d\boldsymbol{\theta}^2}\right] = \mathbf{0}.$$

Note that we can derive $E[\frac{1}{L} \frac{d^2 L}{d\boldsymbol{\theta}^2}]$ as

$$E\left[\frac{1}{L} \frac{d^2 L}{d\boldsymbol{\theta}^2}\right] = E\left[\frac{1}{L} \left(\frac{d^2 L}{d\boldsymbol{\theta}^2}\right)I(\Delta = 1)\right] + E\left[\frac{1}{L} \left(\frac{d^2 L}{d\boldsymbol{\theta}^2}\right)I(\Delta = 0)\right]. \quad (\text{B.2})$$

The first term of the right-hand side of Equation (B.2) can be written as

$$\begin{aligned}
& E\left[\frac{1}{L} \left(\frac{d^2 L}{d\boldsymbol{\theta}^2}\right)I(\Delta = 1)\right] = E\left[\frac{\frac{d^2}{d\boldsymbol{\theta}^2}f(T; \boldsymbol{\theta})}{f(T; \boldsymbol{\theta})}I(T \leq U)\right] \\
& = \int_0^\infty \frac{\frac{d^2}{d\boldsymbol{\theta}^2}f(u; \boldsymbol{\theta})}{f(u; \boldsymbol{\theta})}f(u; \boldsymbol{\theta})G_U(u)du = \frac{d^2}{d\boldsymbol{\theta}^2} \int_0^\infty f(u; \boldsymbol{\theta})G_U(u)du \\
& = \frac{d^2}{d\boldsymbol{\theta}^2}P(T \leq U). \quad (\text{B.3})
\end{aligned}$$

Similarly, the second term of the right-hand side of Equation (B.2) is written as

$$\begin{aligned}
 & E\left[\frac{1}{L}\left(\frac{d^2}{d\boldsymbol{\theta}^2}L\right)I(\Delta=0)\right] = E\left[\frac{\frac{d^2}{d\boldsymbol{\theta}^2}S(U;\boldsymbol{\theta})}{S(U;\boldsymbol{\theta})}I(T>U)\right] \\
 &= \int_0^\infty \frac{\frac{d^2}{d\boldsymbol{\theta}^2}S(u;\boldsymbol{\theta})}{S(u;\boldsymbol{\theta})}S(u;\boldsymbol{\theta})d(1-G_U(u)) = \frac{d^2}{d\boldsymbol{\theta}^2}\int_0^\infty S(u;\boldsymbol{\theta})d(1-G_U(u)) \\
 &= \frac{d^2}{d\boldsymbol{\theta}^2}P(T>U). \tag{B.4}
 \end{aligned}$$

The combination of Equations (B.3) and (B.4) shows that

$$E\left[\frac{1}{L}\frac{d^2L}{d\boldsymbol{\theta}^2}\right] = \frac{d^2}{d\boldsymbol{\theta}^2}[P(T\leq U) + P(T>U)] = 0.$$

APPENDIX C
PROOF OF LEMMA 4.2

We will prove this lemma for $l = 1$. The proof for $l = 2$ is similar and is omitted. Note that

$$\begin{aligned} 0 &\leq E_{\theta_0} \left\{ \int_0^{T_j} u^m h(u; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\} \\ &= - \int_0^\infty \left\{ \int_0^t u^m h(u; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\} d[S_{nj}(t; \boldsymbol{\theta}_0) G_U(t)], \end{aligned} \quad (\text{C.1})$$

where $S_{nj}(t; \boldsymbol{\theta}_0)$ and $G_U(t)$ are survival functions of the survival time T_j and the censoring time U_j , respectively. Note that

$$\begin{aligned} &- \int_0^\infty \left\{ \int_0^t u^m h(u; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\} d[S_{nj}(t; \boldsymbol{\theta}_0) G_U(t)] \\ &= - \left\{ S_{nj}(t; \boldsymbol{\theta}_0) G_U(t) \left[\int_0^t u^m h(u; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right] \right\} \Big|_0^\infty \\ &\quad + \int_0^\infty S_{nj}(t; \boldsymbol{\theta}_0) G_U(t) t^m h(t; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(t)) dt \\ &\leq \int_0^\infty S_{nj}(t; \boldsymbol{\theta}_0) t^m h(t; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(t)) dt. \end{aligned} \quad (\text{C.2})$$

Let

$$\epsilon = \frac{1}{2} \min_{i, g: \lambda_{ig}^{(0)} \neq 0} \{ e^{-B \|\beta_0\|}, |\lambda_{ig}^{(0)}| \},$$

then under the assumption that the j^{th} individual belongs to group g ,

$$\begin{aligned} &\int_0^\infty S_{nj}(t; \boldsymbol{\theta}_0) t^m h(t; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(t)) dt \\ &= \int_0^{\tau_{l-1}} S_{nj}(t; \boldsymbol{\theta}_0) t^m h(t; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(t)) dt + \int_{\tau_{l-1}}^\infty S_{nj}(t; \boldsymbol{\theta}_0) t^m h(t; \boldsymbol{\lambda}', \boldsymbol{\kappa}(j), \mathbf{J}(t)) dt \end{aligned}$$

$$\begin{aligned}
&\leq \tau_{I-1}^m \int_0^{\tau_{I-1}} h(t; \lambda', \kappa(j), \mathbf{J}(t)) dt \\
&\quad + \int_{\tau_{I-1}}^{\infty} \exp\{-e^{\beta_0^T \mathbf{x}_j} \int_0^t h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du\} t^m e^{\lambda'_{I_g} t} dt. \tag{C.3}
\end{aligned}$$

It is observed that the first term in Equation (C.3) is uniformly bounded for λ' satisfying $\|\lambda' - \lambda_0\| < \epsilon$. The second term in Equation (C.3) can be written as

$$\begin{aligned}
&\int_{\tau_{I-1}}^{\infty} \exp\{-e^{\beta_0^T \mathbf{x}_j} \int_0^t h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du\} t^m e^{\lambda'_{I_g} t} dt \\
&\leq \int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \int_{\tau_{I-1}}^t e^{\lambda_{I_g}^{(0)} u} du\} t^m e^{\lambda'_{I_g} t} dt, \tag{C.4}
\end{aligned}$$

since $\|\mathbf{x}_j\| \leq B$ for any $j = 1, 2, \dots, n$ by Assumption 1. We will discuss Equation (C.4) in the following three cases:

Case 1. $\lambda_{I_g}^{(0)} < 0$. $\|\lambda' - \lambda_0\| < \epsilon$ implies $\lambda'_{I_g} < \lambda_{I_g}^{(0)} + \epsilon < \frac{1}{2}\lambda_{I_g}^{(0)} < 0$. Therefore,

$$\int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \int_{\tau_{I-1}}^t e^{\lambda_{I_g}^{(0)} u} du\} t^m e^{\lambda'_{I_g} t} dt \leq \int_{\tau_{I-1}}^{\infty} t^m e^{\frac{1}{2}\lambda_{I_g}^{(0)} t} dt = \Gamma_1 < \infty.$$

Case 2. $\lambda_{I_g}^{(0)} = 0$. The following inequality holds.

$$\int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \int_{\tau_{I-1}}^t e^{\lambda_{I_g}^{(0)} u} du\} t^m e^{\lambda'_{I_g} t} dt \leq \int_{\tau_{I-1}}^{\infty} \exp\{-e^{B\|\beta_0\|} (t - \tau_{I-1})\} t^m e^{\lambda'_{I_g} t} dt.$$

Since $\|\lambda' - \lambda_0\| < \epsilon$ implies that $|\lambda'_{I_g}| = |\lambda'_{I_g} - \lambda_{I_g}^{(0)}| < \epsilon$ and $\lambda'_{I_g} < \frac{1}{2}e^{-B\|\beta_0\|}$,

Equation (C.4) satisfies

$$\begin{aligned}
&\int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \int_{\tau_{I-1}}^t e^{\lambda_{I_g}^{(0)} u} du\} t^m e^{\lambda'_{I_g} t} dt \\
&\leq \int_{\tau_{I-1}}^{\infty} \exp\{-e^{B\|\beta_0\|} (\frac{1}{2}t - \tau_{I-1})\} t^m dt = \Gamma_2 < \infty.
\end{aligned}$$

Case 3. $\lambda_{I_g}^{(0)} > 0$. Equation (C.4) can be written as

$$\begin{aligned}
&\int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \int_{\tau_{I-1}}^t e^{\lambda_{I_g}^{(0)} u} du\} t^m e^{\lambda'_{I_g} t} dt \\
&= \int_{\tau_{I-1}}^{\infty} \exp\{-e^{-B\|\beta_0\|} \frac{e^{\lambda_{I_g}^{(0)} t} - e^{\lambda_{I_g}^{(0)} \tau_{I-1}}}{\lambda_{I_g}^{(0)}}\} t^m e^{\lambda'_{I_g} t} dt \leq \Gamma_3 < \infty,
\end{aligned}$$

since λ'_{tg} is uniformly bounded. Hence, there exists a $\Gamma > 0$ such that

$$E_{\theta_0} \left\{ \int_0^{T_j^*} u^m h(u; \lambda', \kappa(j), \mathbf{J}(u)) du \right\} \leq \Gamma,$$

for λ' satisfying $\|\lambda' - \lambda_0\| < \epsilon$.

APPENDIX D
PROOF OF LEMMA 4.3

Note that for $0 < \delta < \infty$,

$$\begin{aligned}
& E_{\theta_0} \left\{ \int_0^{T_j^*} u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \right\} \\
&= - \int_0^\infty \left[\int_0^t u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \right] d[S_{nj}(t; \theta_0) G_U(t)] \\
&\geq - \int_\delta^\infty \left[\int_0^\delta u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \right] d[S_{nj}(t; \theta_0) G_U(t)] \\
&= - \int_0^\delta u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \int_\delta^\infty d[S_{nj}(t; \theta_0) G_U(t)] \\
&= \left[\int_0^\delta u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \right] [S_{nj}(\delta; \theta_0) G_U(\delta) - S_{nj}(\infty; \theta_0) G_U(\infty)].
\end{aligned}$$

We observe that

$$\begin{aligned}
& S_{nj}(\delta; \theta_0) G_U(\delta) - S_{nj}(\infty; \theta_0) G_U(\infty) \geq S_{nj}(\delta; \theta_0) [G_U(\delta) - G_U(\infty)] \\
&\geq \exp\{-e^{B\|\beta_0\|} \int_0^\delta h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du\} [G_U(\delta) - G_U(\infty)] \\
&\rightarrow 1 - G_U(\infty),
\end{aligned}$$

as $\delta \rightarrow 0$ uniformly for $j = 1, 2, \dots, n$. Therefore, we can find a $\delta_0 > 0$ and an $\epsilon > 0$ such that

$$S_{nj}(\delta_0; \theta_0) G_U(\delta_0) - S_{nj}(\infty; \theta_0) G_U(\infty) > \epsilon$$

and

$$\int_0^{\delta_0} u^m h^l(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du > \epsilon.$$

Thus the expectation term in Equation (4.1) $> \epsilon^2$ for all $j = 1, 2, \dots, n$. If we take $\gamma = \epsilon^2$, the lemma is proved.

APPENDIX E
PROOF OF LEMMA 4.4

Let $\mathbf{c}^T = (\mathbf{a}^T, \mathbf{b}^T)$ be a $(p + IG)$ -dimensional unit vector, where $\mathbf{a}$ is p -dimensional and $\mathbf{b}^T = (b_{11}, \dots, b_{I1}, \dots, b_{IG}, \dots, b_{IG})^T$ is IG -dimensional. It suffices to show that

$$\mathbf{c}^T D_n \mathbf{c} \geq n\gamma,$$

for any unit vector $\mathbf{c}$. We showed in Chapter 3 that

$$\mathbf{c}^T D_n \mathbf{c} = \sum_{j=1}^n e^{\beta^T \mathbf{x}_j} E \int_0^{T_j^*} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) (\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j))^2 du.$$

By expanding the expectation in the above equation, we obtain

$$\begin{aligned} & \mathbf{c}^T D_n \mathbf{c} \\ &= \sum_{j=1}^n e^{\beta^T \mathbf{x}_j} \int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) (\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j))^2 du \right\} dF_{T_j^*}(t) \\ &= - \sum_{j=1}^n e^{\beta^T \mathbf{x}_j} \int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) (\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j))^2 du \right\} d[S_j(t)G_U(t)]. \end{aligned}$$

Since $\|\mathbf{x}_j\| \leq B$ by Assumption 1,

$$\begin{aligned} & \mathbf{c}^T D_n \mathbf{c} \\ &\geq - \sum_{j=1}^n e^{-B\|\boldsymbol{\beta}\|} \int_0^\infty \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) (\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j))^2 du d[S_j(t)G_U(t)] \\ &= e^{-B\|\boldsymbol{\beta}\|} \sum_{g=1}^G \sum_{j \in \text{group } g} - \int_0^\infty \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) (\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j))^2 du \\ & \quad d[S_j(t)G_U(t)]. \end{aligned}$$

(E.1)

Assume that the j^{th} individual belongs to the treatment group g , then

$$u\mathbf{b}^T\mathbf{Q}(u, j) = u \sum_{i=1}^I b_{ig} I_{\{u \in [\tau_{i-1}, \tau_i]\}}$$

and

$$[u\mathbf{b}^T\mathbf{Q}(u, j)]^2 = u^2 \sum_{i=1}^I b_{ig}^2 I_{\{u \in [\tau_{i-1}, \tau_i]\}}.$$

Note that there exist a g_0 where $1 \leq g_0 \leq G$ and a i_0 where $1 \leq i_0 \leq I$, such that

$$b_{i_0 g_0}^2 \geq \frac{\|\mathbf{b}\|^2}{IG} = \frac{1 - \|\mathbf{a}\|^2}{IG}. \quad (\text{E.2})$$

We will discuss the behavior of the right-hand side of Equation (E.1) in the following four cases.

Case 1. $\|\mathbf{a}\|^2 \leq \frac{\tau_1^2}{2(8B^2GI + \tau_1^2)}$ and $i_0 = 1$. Note that for each j in treatment group g_0 ,

$$\begin{aligned} & - \int_0^\infty \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u\mathbf{b}^T \mathbf{Q}(u, j)]^2 du d[S_j(t)G_U(t)] \\ \geq & - \int_{\tau_1}^\infty \int_{\frac{1}{2}\tau_1}^{\tau_1} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u\mathbf{b}^T \mathbf{Q}(u, j)]^2 du d[S_j(t)G_U(t)]. \end{aligned}$$

Because $u \geq \frac{1}{2}\tau_1$, $\|\mathbf{x}\| \leq B$, and $b_{i_0 g_0}^2 \geq \frac{1 - \|\mathbf{a}\|^2}{IG}$, we observe that

$$\begin{aligned} & [\mathbf{a}^T \mathbf{x}_j + u\mathbf{b}^T \mathbf{Q}(u, j)]^2 \geq \frac{1}{2} [u\mathbf{b}^T \mathbf{Q}(u, j)]^2 - (\mathbf{a}^T \mathbf{x}_j)^2 \\ \geq & \frac{1}{2} u^2 b_{i_0 g_0}^2 - (\mathbf{a}^T \mathbf{x}_j)^2 \geq \frac{1}{8} \tau_1^2 b_{i_0 g_0}^2 - (\mathbf{a}^T \mathbf{x}_j)^2 \\ \geq & \frac{\tau_1^2}{8IG} (1 - \|\mathbf{a}\|^2) - B^2 \|\mathbf{a}\|^2 \geq \frac{\tau_1^2}{16IG}. \end{aligned}$$

Hence,

$$\begin{aligned} & - \int_0^\infty \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u\mathbf{b}^T \mathbf{Q}(u, j)]^2 du d[S_j(t)G_U(t)] \\ \geq & - \frac{\tau_1^2}{16GI} \int_{\tau_1}^\infty \left[\int_{\frac{1}{2}\tau_1}^{\tau_1} e^{\lambda_{i_0 g_0} u} du \right] d[S_j(t)G_U(t)] \\ = & \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{i_0 g_0}} (e^{\lambda_{i_0 g_0} \tau_1} - e^{\frac{1}{2}\lambda_{i_0 g_0} \tau_1}) [S_j(\tau_1)G_U(\tau_1) - S_j(\infty)G_U(\infty)] \end{aligned}$$

$$\begin{aligned}
&\geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{I_{g_0}}} (e^{\lambda_{I_{g_0}} \tau_1} - e^{\frac{1}{2} \lambda_{I_{g_0}} \tau_1}) S_j(\tau_1) [G_U(\tau_1) - G_U(\infty)] \\
&\geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{I_{g_0}}} (e^{\lambda_{I_{g_0}} \tau_1} - e^{\frac{1}{2} \lambda_{I_{g_0}} \tau_1}) \exp\{-e^{B\|\beta\|} \int_0^{\tau_1} e^{\lambda_{I_{g_0}} u} du\} [G_U(\tau_1) - G_U(\infty)] \\
&= \gamma_1 > 0,
\end{aligned}$$

by Assumption 4.

Case 2. $\|\mathbf{a}\|^2 \leq \frac{\tau_1^2}{2(8B^2GI + \tau_1^2)}$ and $i_0 = I$. For each j in treatment group g_0 ,

$$\begin{aligned}
&- \int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t) G_U(t)] \\
&\geq - \int_{\tau_{I-1}}^\infty \left\{ \int_{\tau_{I-1}}^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t) G_U(t)] \\
&\geq - \int_{\tau_{I-1} + \delta}^\infty \left\{ \int_{\tau_{I-1}}^{\tau_{I-1} + \delta} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t) G_U(t)].
\end{aligned}$$

Since $u \geq \tau_{I-1} \geq \frac{1}{2} \tau_1$ in the above expression, with the additional conditions that $b_{I_{g_0}}^2 \geq \frac{1 - \|\mathbf{a}\|^2}{IG}$ and $\|\mathbf{x}_j\| \leq B$ for any $j = 1, 2, \dots, n$, we obtain that

$$[\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 \geq \frac{\tau_1^2}{16IG}.$$

Therefore,

$$\begin{aligned}
&- \int_{\tau_{I-1} + \delta}^\infty \left\{ \int_{\tau_{I-1}}^{\tau_{I-1} + \delta} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t) G_U(t)] \\
&\geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{I_{g_0}}} (e^{\lambda_{I_{g_0}}(\tau_{I-1} + \delta)} - e^{\lambda_{I_{g_0}} \tau_{I-1}}) [S_j(\tau_{I-1} + \delta) G_U(\tau_{I-1} + \delta) - S_j(\infty) G_U(\infty)] \\
&\geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{I_{g_0}}} (e^{\lambda_{I_{g_0}}(\tau_{I-1} + \delta)} - e^{\lambda_{I_{g_0}} \tau_{I-1}}) S_j(\tau_{I-1} + \delta) [G_U(\tau_{I-1} + \delta) - G_U(\infty)].
\end{aligned} \tag{E.3}$$

Note that

$$\begin{aligned}
&S_j(\tau_{I-1} + \delta) [G_U(\tau_{I-1} + \delta) - G_U(\infty)] \\
&\xrightarrow{\delta \rightarrow 0} S_j(\tau_{I-1}) [G_U(\tau_{I-1}) - G_U(\infty)] \\
&\geq \exp\{-e^{B\|\beta\|} \int_0^{\tau_{I-1}} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du\} [G_U(\tau_{I-1}) - G_U(\infty)] > \epsilon > 0,
\end{aligned}$$

where Assumption 4 is applied. Therefore, there exists a $\delta > 0$ such that

$$-\int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t)G_U(t)] > \gamma_2 > 0.$$

Case 3. $\|\mathbf{a}\|^2 \leq \frac{\tau_1^2}{2(8B^2GI + \tau_1^2)}$ and $1 < i_0 < I$. Note that for $u \in [\tau_{i_0-1}, \tau_{i_0})$ and for each j in treatment group g_0 , the following inequality also holds:

$$[\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 \geq \frac{\tau_1^2}{16IG}.$$

Therefore,

$$\begin{aligned} & -\int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du \right\} d[S_j(t)G_U(t)] \\ & \geq -\int_{\tau_{i_0}}^\infty \int_{\tau_{i_0-1}}^{\tau_{i_0}} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) [\mathbf{a}^T \mathbf{x}_j + u \mathbf{b}^T \mathbf{Q}(u, j)]^2 du d[S_j(t)G_U(t)] \\ & \geq \frac{\tau_1^2}{16GI} \int_{\tau_{i_0-1}}^{\tau_{i_0}} e^{\lambda_{i_0g_0}u} du [S_j(\tau_{i_0})G_U(\tau_{i_0}) - S_j(\infty)G_U(\infty)] \\ & \geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{i_0g_0}} (e^{\lambda_{i_0g_0}\tau_{i_0}} - e^{\lambda_{i_0g_0}\tau_{i_0-1}}) S_j(\tau_{i_0}) [G_U(\tau_{i_0}) - G_U(\infty)] \\ & \geq \frac{\tau_1^2}{16GI} \frac{1}{\lambda_{i_0g_0}} (e^{\lambda_{i_0g_0}\tau_{i_0}} - e^{\lambda_{i_0g_0}\tau_{i_0-1}}) \exp\{-e^{B\|\boldsymbol{\beta}\|} \int_0^{\tau_{i_0}} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du\} \\ & \quad [G_U(\tau_{i_0}) - G_U(\infty)] \\ & = \gamma_3 > 0, \end{aligned}$$

where Assumption 4 is used again.

By combining the above three cases, we obtain that

$$\mathbf{c}^T D_n \mathbf{c} \geq e^{-B\|\boldsymbol{\beta}\|} n_{g_0} \gamma',$$

where $\gamma' = \min(\gamma_1, \gamma_2, \gamma_3)$ and n_{g_0} is the number of individuals in group g_0 . Under Assumption 3 the following holds:

$$\min_{1 \leq j \leq n} \frac{n_j}{n} > \delta_1,$$

when n is sufficiently large. Therefore,

$$\mathbf{c}^T D_n \mathbf{c} \geq e^{-B\|\beta\|} n \delta_1 \gamma' = n\gamma$$

where $\gamma = \delta_1 \gamma' e^{-B\|\beta\|}$.

Case 4. $\|\mathbf{a}\|^2 > \frac{\tau_1^2}{2(8B^2GI + \tau_1^2)} \stackrel{\text{say}}{=} \epsilon_1^2$. Note that

$$|\mathbf{u} \mathbf{b}^T \mathbf{Q}(u, j)| \leq |u|GI.$$

Therefore,

$$\begin{aligned} (\mathbf{a}^T \mathbf{x}_j + \mathbf{u} \mathbf{b}^T \mathbf{Q}(u, j))^2 &\geq \frac{1}{2} (\mathbf{a}^T \mathbf{x}_j)^2 - [\mathbf{u} \mathbf{b}^T \mathbf{Q}(u, j)]^2 \\ &\geq \frac{1}{2} (\mathbf{a}^T \mathbf{x}_j)^2 - u^2 G^2 I^2. \end{aligned}$$

By Assumption 2, there exist an $\epsilon > 0$, a $\delta > 0$, and a $n_0 > 0$. When $n > n_0$, there exists a subset A of $\{1, 2, \dots, n\}$ with size $n_A > n\delta$ such that

$$|\frac{\mathbf{a}^T}{\|\mathbf{a}\|} \mathbf{x}_j| > \epsilon \text{ for } j \in A. \quad (\text{E.4})$$

Now we assume $n > n_0$ and (E.4) is satisfied. Therefore, if we denote $\frac{\epsilon \epsilon_1}{2GI}$ by δ_1 , for $u \in (0, \delta_1]$,

$$\begin{aligned} [\mathbf{a}^T \mathbf{x}_j + \mathbf{u} \mathbf{b}^T \mathbf{Q}(u, j)]^2 &\geq \frac{1}{2} (\mathbf{a}^T \mathbf{x}_j)^2 - u^2 G^2 I^2 \\ &\geq \frac{1}{2} \epsilon^2 \|\mathbf{a}\|^2 - \delta_1^2 G^2 I^2 \geq \frac{1}{2} \epsilon^2 \epsilon_1^2 - \frac{\epsilon^2 \epsilon_1^2}{4G^2 I^2} G^2 I^2 \\ &= \frac{1}{4} \epsilon^2 \epsilon_1^2. \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbf{c}^T D_n \mathbf{c} \\ &\geq e^{-B\|\beta\|} \sum_{j \in A} - \int_0^\infty \left\{ \int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \frac{\epsilon^2 \epsilon_1^2}{4} I_{[u \in (0, \delta_1)]} du \right\} d[S_j(t) G_U(t)]. \end{aligned}$$

Without loss of generality, let us assume $\delta_1 < \tau_1$. For the j^{th} individual in treatment group g ,

$$\begin{aligned}
& - \int_0^\infty \left[\int_0^t h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \frac{\epsilon^2 \epsilon_1^2}{4} I_{[u \in (0, \delta_1)]} du \right] d[S_j(t)G_U(t)] \\
& \geq - \int_{\tau_1}^\infty \left[\int_0^{\delta_1} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \frac{\epsilon^2 \epsilon_1^2}{4} du \right] d[S_j(t)G_U(t)] \\
& \geq \left[\int_0^{\delta_1} e^{\lambda_{1g} u} du \right] \frac{\epsilon^2 \epsilon_1^2}{4} [S_j(\tau_1)G_U(\tau_1) - S_j(\infty)G_U(\infty)] > \gamma'_g > 0,
\end{aligned}$$

where Assumption 4 is used. Note that γ'_g depends on g . If we take

$$\gamma_0 = \min_{1 \leq g \leq G} \gamma'_g > 0 \text{ and}$$

denote $\delta \gamma_0 e^{-B\|\boldsymbol{\beta}\|}$ by γ , then

$$\mathbf{c}^T D_n \mathbf{c} \geq e^{-B\|\boldsymbol{\beta}\|} n_A \gamma_0 > e^{-B\|\boldsymbol{\beta}\|} n \delta \gamma_0 = n \gamma.$$

APPENDIX F
PROOF OF LEMMA 4.5

From Lemma 4.1, we observe that

$$E[\mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}_0) - D_n] D_n^{-1/2} \mathbf{c}] = \mathbf{c}^T D_n^{-1/2} [D_n - D_n] D_n^{-1/2} \mathbf{c} = 0.$$

In addition we have

$$\begin{aligned} & \mathbf{c}^T D_n^{-1/2} F_n(\boldsymbol{\theta}_0) D_n^{-1/2} \mathbf{c} \\ &= \sum_{j=1}^n e^{\beta^T \mathbf{x}_j} \int_0^{T_j^*} h(u; \boldsymbol{\lambda}_0, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} \begin{pmatrix} \mathbf{x}_j^T & u \mathbf{Q}^T(u, j) \end{pmatrix} \\ & \quad D_n^{-1/2} \mathbf{c} du. \end{aligned}$$

We showed in Lemma 4.4 that $\lambda_{\min}(D_n) \geq n\gamma$. This implies that $\lambda_{\max}(D_n^{-1/2}) \leq \frac{1}{\sqrt{n\gamma}}$. Therefore,

$$\begin{aligned} & \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} \begin{pmatrix} \mathbf{x}_j^T & u \mathbf{Q}^T(u, j) \end{pmatrix} D_n^{-1/2} \mathbf{c} = \left[\mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} \right]^2 \\ & \leq \begin{pmatrix} \mathbf{x}_j^T & u \mathbf{Q}^T(u, j) \end{pmatrix} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} \lambda_{\max}^2(D_n^{-1/2}) \leq \frac{1}{n\gamma} (B^2 + IG u^2). \end{aligned} \quad (\text{F.1})$$

Since T_j^* , $j = 1, 2, \dots, n$ are independent,

$$\begin{aligned} & V \left\{ \mathbf{c}^T D_n^{-1/2} (F_n(\boldsymbol{\theta}_0) - D_n) D_n^{-1/2} \mathbf{c} \right\} \\ &= \sum_{j=1}^n e^{2\beta_0^T \mathbf{x}_j} V \left\{ \int_0^{T_j^*} h(u; \boldsymbol{\lambda}_0, \boldsymbol{\kappa}(j), \mathbf{J}(u)) \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} \begin{pmatrix} \mathbf{x}_j^T & u \mathbf{Q}^T(u, j) \end{pmatrix} \right. \\ & \quad \left. D_n^{-1/2} \mathbf{c} du \right\} \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{j=1}^n e^{2B\|\beta_0\|} E \left\{ \int_0^{T_j^*} h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(u, j) \end{pmatrix} (\mathbf{x}_j^T u \mathbf{Q}^T(u, j)) \right. \\
&\quad \left. D_n^{-1/2} \mathbf{c} du \right\}^2 \\
&\leq e^{2B\|\beta_0\|} \sum_{j=1}^n E \left\{ \int_0^{T_j^*} h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) \frac{1}{n\gamma} (B^2 + I G u^2) du \right\}^2 \\
&\leq e^{2B\|\beta_0\|} \frac{(B^2 + I G)^2}{n^2 \gamma^2} \sum_{j=1}^n \max_{m=0,2} E \left\{ \int_0^{T_j^*} u^m h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \right\}^2 \\
&\leq e^{2B\|\beta_0\|} \frac{(B^2 + I G)^2}{n^2 \gamma^2} n \Gamma \\
&= O(n^{-1}).
\end{aligned}$$

The last inequality holds by Lemma 4.2 and Γ is a constant depending on λ_0 .

Therefore, for any $\epsilon > 0$,

$$P(|\mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}_0) - D_n] D_n^{-1/2} \mathbf{c}| > \epsilon) \leq \frac{V \left\{ \mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}_0) - D_n] D_n^{-1/2} \mathbf{c} \right\}}{\epsilon^2} \rightarrow 0,$$

as $n \rightarrow \infty$ by the Chebyshev's inequality. This implies

$$\mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}_0) - D_n] D_n^{-1/2} \mathbf{c} \xrightarrow{P} 0.$$

APPENDIX G
PROOF OF LEMMA 4.6

Let $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)$ and $\boldsymbol{\theta}_0^T = (\boldsymbol{\beta}_0^T, \boldsymbol{\lambda}_0^T)$, then

$$\begin{aligned} & \mathbf{c}^T D_n^{-1/2} [F_n(\boldsymbol{\theta}) - F_n(\boldsymbol{\theta}_0)] D_n^{-1/2} \mathbf{c} \\ &= \sum_{j=1}^n \int_0^{T_j^*} [e^{\boldsymbol{\beta}^T \mathbf{x}_j} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) - e^{\boldsymbol{\beta}_0^T \mathbf{x}_j} h(u; \boldsymbol{\lambda}_0, \boldsymbol{\kappa}(j), \mathbf{J}(u))] \\ & \quad \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u \mathbf{Q}(j) \end{pmatrix} (\mathbf{x}^T u \mathbf{Q}^T(j)) D_n^{-1/2} \mathbf{c} du. \end{aligned} \quad (\text{G.1})$$

By Taylor's expansion,

$$\begin{aligned} & [e^{\boldsymbol{\beta}^T \mathbf{x}_j} h(u; \boldsymbol{\lambda}, \boldsymbol{\kappa}(j), \mathbf{J}(u)) - e^{\boldsymbol{\beta}_0^T \mathbf{x}_j} h(u; \boldsymbol{\lambda}_0, \boldsymbol{\kappa}(j), \mathbf{J}(u))] \\ &= [\mathbf{x}_j^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + u \mathbf{1}^T (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)] e^{\tilde{\boldsymbol{\beta}}_j^T \mathbf{x}_j} h(u; \tilde{\boldsymbol{\lambda}}_j, \boldsymbol{\kappa}(j), \mathbf{J}(u)), \end{aligned}$$

where $\mathbf{1}$ is a IG -dimensional vector of 1 and $(\tilde{\boldsymbol{\beta}}_j^T, \tilde{\boldsymbol{\lambda}}_j^T) = \alpha_j(\boldsymbol{\beta}_0^T, \boldsymbol{\lambda}_0^T) + (1 - \alpha_j)(\boldsymbol{\beta}_n^T, \boldsymbol{\lambda}_n^T)$ for $0 < \alpha_j < 1$. By Equation (4.3) and Assumption 1,

$$|\mathbf{x}_j^T (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + u \mathbf{1}^T (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0)| \leq (B + IG u) \frac{1}{\sqrt{n}} \sqrt{\frac{A}{\gamma}},$$

which, together with (F.1), implies that the absolute value of the integrand in (G.1)

$$\leq (B + IG u) \frac{1}{\sqrt{n}} \sqrt{\frac{A}{\gamma}} e^{B \|\boldsymbol{\beta}_0\|} \frac{1}{n\gamma} (B^2 + IG u^2) h(u; \tilde{\boldsymbol{\lambda}}_j, \boldsymbol{\kappa}(j), \mathbf{J}(u)). \quad (\text{G.2})$$

By Lemma 4.2,

$$\begin{aligned} & |E[\mathbf{c}^T D_n^{-1/2} (F_n(\boldsymbol{\theta}) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}]| \\ &\leq \sum_{j=1}^n E_{\boldsymbol{\theta}_0} \left\{ \int_0^{T_j^*} (B + IG u) \frac{1}{\sqrt{n}} \sqrt{\frac{A}{\gamma}} e^{B \|\boldsymbol{\beta}_0\|} \frac{1}{n\gamma} (B^2 + IG u^2) h(u; \tilde{\boldsymbol{\lambda}}_j, \boldsymbol{\kappa}(j), \mathbf{J}(u)) du \right\} \end{aligned}$$

$$\begin{aligned}
&\leq n(B + IG)(B^2 + IG)e^{B\|\beta_0\|} \frac{\sqrt{A}}{n\sqrt{n}\gamma\sqrt{\gamma}} \max_{0 \leq m \leq 3} E \int_0^{T_j^*} u^m h(u; \tilde{\lambda}_j, \kappa(j), \mathbf{J}(u)) du \\
&\leq n(B + IG)(B^2 + IG)e^{B\|\beta_0\|} \frac{\sqrt{A}}{n\sqrt{n}\gamma\sqrt{\gamma}} \Gamma \\
&= O(n^{-1/2}).
\end{aligned}$$

Furthermore, by the independence of T_j^* s,

$$\begin{aligned}
&V[\mathbf{c}^T D_n^{-1/2}(F_n(\boldsymbol{\theta}) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}] \\
&= \sum_{j=1}^n V\left[\int_0^{T_j^*} [e^{\beta^T \mathbf{x}_j} h(u; \lambda, \kappa(j), \mathbf{J}(u)) - e^{\beta_0^T \mathbf{x}_j} h(u; \lambda_0, \kappa(j), \mathbf{J}(u))] \right. \\
&\quad \left. \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u\mathbf{Q}(j) \end{pmatrix} (\mathbf{x}^T u\mathbf{Q}^T(j)) D_n^{-1/2} \mathbf{c} du\right],
\end{aligned}$$

where the integrand in the above equation is the same as in (G.1). Hence, from Equation (G.2) and Lemma 4.2,

$$\begin{aligned}
&V[\mathbf{c}^T D_n^{-1/2}(F_n(\boldsymbol{\theta}) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}] \\
&\leq \sum_{j=1}^n E\left[\int_0^{T_j^*} [e^{\beta^T \mathbf{x}_j} h(u; \lambda, \kappa(j), \mathbf{J}(u)) - e^{\beta_0^T \mathbf{x}_j} h(u; \lambda_0, \kappa(j), \mathbf{J}(u))] \right. \\
&\quad \left. \mathbf{c}^T D_n^{-1/2} \begin{pmatrix} \mathbf{x}_j \\ u\mathbf{Q}(j) \end{pmatrix} (\mathbf{x}^T u\mathbf{Q}^T(j)) D_n^{-1/2} \mathbf{c} du\right]^2 \\
&\leq (B + IG)^2 (B^2 + IG)^2 \left(\frac{\sqrt{A}}{n\sqrt{n}\gamma\sqrt{\gamma}}\right)^2 \sum_{j=1}^n \max_{0 \leq m \leq 3} E\left[\int_0^{T_j^*} u^m h(u; \tilde{\lambda}_j, \kappa(j), \mathbf{J}(u)) du\right]^2 \\
&\leq (B + IG)^2 (B^2 + IG)^2 \frac{A}{n^3 \gamma^3} n\Gamma \\
&= O(n^{-2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&E[\mathbf{c}^T D_n^{-1/2}(F_n(\boldsymbol{\theta}_n) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}]^2 \\
&= V[\mathbf{c}^T D_n^{-1/2}(F_n(\boldsymbol{\theta}_n) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}] + [E\mathbf{c}^T D_n^{-1/2}(F_n(\boldsymbol{\theta}_n) - F_n(\boldsymbol{\theta}_0)) D_n^{-1/2} \mathbf{c}]^2 \\
&= O(n^{-1}).
\end{aligned}$$

Hence, the lemma is proved since mean square convergence implies convergence in probability.

APPENDIX H
PROOF OF LEMMA 4.7

From Lemmas 4.5 and 4.6,

$$\mathbf{c}^T D_n^{-1/2} (F_n(\boldsymbol{\theta}) - D_n) D_n^{-1/2} \mathbf{c} \xrightarrow{p} 0, \quad (\text{H.1})$$

where $\mathbf{c}$ is a $(p + IG)$ -dimensional unit vector. As a result of Equation (H.1),

$$\mathbf{c}^T D_n^{-1/2} F_n(\boldsymbol{\theta}) D_n^{-1/2} \mathbf{c} \xrightarrow{p} 1,$$

which implies that

$$\lim_{n \rightarrow \infty} \lambda_{\max}(D_n^{-1/2} F(\boldsymbol{\theta}) D_n^{-1/2}) = \lim_{n \rightarrow \infty} \lambda_{\min}(D_n^{-1/2} F(\boldsymbol{\theta}) D_n^{-1/2}) = 1$$

in probability. Therefore,

$$D_n^{-1/2} F(\boldsymbol{\theta}) D_n^{-1/2} \xrightarrow{p} \mathbf{I}_{(p+IG) \times (p+IG)}.$$

APPENDIX I
PROOF OF LEMMA 4.8

It suffices to show that for each k , $k = 1, 2, \dots, p + IG$,

$$E_{\theta_0} \|S_{njk}(\boldsymbol{\theta})\|^l$$

is uniformly bounded for $j = 1, 2, \dots, n$, where $S_{njk}(\boldsymbol{\theta})$ is the k^{th} component of the vector $S_{nj}(\boldsymbol{\theta})$. From Equations (3.2) and (3.3), we observe that $|S_{njk}(\boldsymbol{\theta})|$ is less than or equal to the maximum of the following two quantities:

$$\begin{aligned} B + Be^{B\|\beta_0\|} \int_0^{T_j^*} h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du \\ T_j + e^{B\|\beta_0\|} \int_0^{T_j^*} h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du. \end{aligned}$$

Note that T_j is the survival time of the j^{th} individual and $T_j^* = \min(T_j, U_j)$, where U_j is the censoring time associated with the j^{th} individual. To show $E_{\theta_0} |S_{njk}|^l$ is bounded uniformly for j , it suffices to show that $E[T_j]^l$ is uniformly bounded since we have shown the uniform boundedness of $E_{\theta_0} \int_0^{T_j^*} h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) du$ in Lemma 4.2. Assume that the j^{th} individual belongs to the treatment group g , then

$$\begin{aligned} E[T_j]^l &= \int_0^\infty t^l h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) e^{\beta_0^T \mathbf{x}_j} S_{nj}(t; \boldsymbol{\theta}_0) dt \\ &\leq e^{B\|\beta_0\|} \tau_{I-1}^l \int_0^{\tau_{I-1}} h(t; \lambda_0, \kappa(j), \mathbf{J}(u)) dt \\ &\quad + e^{B\|\beta_0\|} \int_{\tau_{I-1}}^\infty t^l h(u; \lambda_0, \kappa(j), \mathbf{J}(u)) S_{nj}(t; \boldsymbol{\theta}_0) dt. \end{aligned} \quad (\text{I.1})$$

Note that the first term in Equation (I.1) is uniformly bounded for any j . The second term is the same as in Equation (C.4) where the parameters are replaced by their true value. Hence, the uniform boundedness is proved.

APPENDIX J SIMPSON'S FORMULA

Suppose we want to approximate one integral by dividing the integration interval to N_0 points. Assume a is the lower bound of the integral, b is the upper bound of the integral, and the integration form we are interested is

$$\int_a^b f(c)dc, \quad (\text{J.1})$$

then we will divide the interval (a, b) into N_0 intervals with the end points of the intervals represented as $\{x_j\}_{j=0}^{N_0}$, where $x_0 = a$, $x_{N_0} = b$, $x_j = a + jh$, and $h = \frac{b-a}{N_0}$.

The numerical approximation of (J.1) is then written as

$$\int_a^b f(c)dc = \frac{h}{6} \left\{ f(x_0) + f(x_{N_0}) + 2 \sum_{j=1}^{N_0-1} f(x_j) + 4 \sum_{j=0}^{N_0-1} f\left(\frac{x_j + x_{j+1}}{2}\right) \right\}.$$

REFERENCES

- Aitken, M., Anderson, D., Francis, B. and Hinde, J. (1989). *Statistical modelling in GLIM*, Clarendon Press [Oxford University Press].
- Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment, *Journal of the American Statistical Association* **47**: 501–515.
- Bithell, J. F. and Upton, R. G. (1977). A mixed model for survival applied to British children with neuroblastoma, *Recent Developments in Statistics, Proceedings of the 1976 European Meeting of Statisticians*, North-Holland/Elsevier (Amsterdam), pp. 635–646.
- Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy, *Journal of the Royal Statistical Society* **11**: 15–53.
- Cantor, A. B. and Shuster, J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data, *Statistics in Medicine* **11**: 931–937.
- Chen, M., Ibrahim, J. and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**: 909–949.
- Doornik, J. A. (1998). *Ox - An Object-Oriented Matrix Programming Language*, Timberlake Consultants (London).
- Farewell, V. T. (1977). A model for a binary variable with time-censored observations, *Biometrika* **64**: 43–46.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* **38**: 1041–1046.
- Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk?, *The Canadian Journal of Statistics* **14**: 257–262.
- Friedman, M. (1981). Piecewise exponential models for survival data with covariates, *Annals of Statistics* **10**: 101–113.
- Gieser, P. W., Chang, M. N., Rao, P. V., Shuster, J. J. and Pullen, J. (1998). Modelling cure rates using the Gompertz model with covariate information, *Statistics in Medicine* **17**: 831–839.

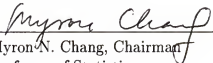
- Goldman, A. I. (1984). Survivorship analysis when cure is a possibility: A Monte Carlo study, *Statistics in Medicine* **3**: 153–163.
- Gordon, N. H. (1990). Application of the theory of finite mixtures for the estimation of "cure" rates of treated cancer patients, *Statistics in Medicine* **9**: 397–407.
- Gray, R. J. and Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates, *Biometrics* **45**: 899–904.
- Halpern, J. and Brown, B. W. J. (1987). Cure rate models: Power of the logrank and generalized Wilcoxon tests, *Statistics in Medicine* **6**: 483–489.
- Jung, S.-H. (1996). Regression analysis for long-term survival rate, *Biometrika* **83**: 227–232.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley (New York).
- Karrison, T. (1987). Restricted mean life with adjustment for covariates, *Journal of the American Statistical Association* **82**: 1169–1176.
- Kuk, A. Y. C. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression, *Biometrika* **79**: 531–541.
- Laska, E. M. and Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model, *Biometrics* **48**: 1223–1234.
- Maller, R. A. (1988). On the exponential model for survival, *Biometrika* **75**: 582–586.
- Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*, Wiley (New York).
- Peng, Y., Dear, K. B. G. and Denham, J. W. (1998). A generalized f mixture model for cure rate estimation, *Statistics in Medicine* **17**: 813–830.
- Rao, C. (1973). *Linear Statistical Inference and It's Applications (Second Edit ion)*, Wiley (New York).
- Ravindranath, Y., Yeager, A. M., Chang, M. N., Steuber, C. P., Krischer, J., Graham-Pole, J., Carroll, A., Inoue, S., Camitta, B. and Weinstein, H. J. (1996). Autologous bone marrow transplantation vesus intensive consolidation chemotherapy for acute myeloid leukemia in childhood, *The New England Journal of Medicine* **334**: 1428–1434.
- Risch, N. (1983). Estimating morbidity risks with variable age of onset: Review of methods and a maximum likelihood approach, *Biometrics* **39**: 929–939.

- Sposto, R., Sather, H. N. and Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients who are cured, *Biometrics* **48**: 87–99.
- Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distribution, *Biometrika* **64**: 156–160.
- Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models, *Biometrics* **51**: 899–907.
- Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors, *Biometrics* **54**: 1508–1516.
- Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of “permanent employment” in Japan, *Journal of the American Statistical Association* **87**: 284–292.

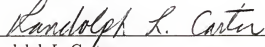
BIOGRAPHICAL SKETCH

Haoyi Chen was born on November 29, 1972, in Jinan, People's Republic of China. She received her Bachelor of Economics degree from Shanghai University of Finance and Economics in July 1993, majoring in statistics. After two years of industrial experience in business statistics in China, she joined the graduate program in the Department of Statistics at the University of Florida. She received her Master of Statistics degree in January 1998. She is expected to receive her PhD degree in December 2001.

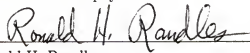
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Myron N. Chang, Chairman
Professor of Statistics

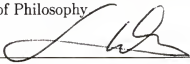
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Randolph L. Carter
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Ronald H. Randles
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Samuel S. Wu
Assistance Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Michael B. Resnick
Professor of Pediatrics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 2001

Dean, Graduate School